Co-funded by
the European Union

EducationalAI

Ethical & Legal Guide

on the responsible integration of AI in education

August 2025

Ethical & Legal Guide

on the responsible integration of AI in education

Project Title	Empowering Educational Leaders with AI Strategies
Project Acronym	EducationalAI
Project Number	2024-1-FI01-KA220-ADU-000255928
Dissemination Level	Public
Delivery Date	2025-08
Author(s)	Maria Bertel, Evelyne Putz, Julian Priebsch, Universität Graz; Dovilė Balcevičienė, Danguolė Mačiulienė, Ričardas Balcevičius, Vilnius Ozas Gymnasium; Cristina Obae, Sykli
Contributors	Academia de Studii Economice din Bucuresti, ARTKIT Euforija, Hogskulen pa Vestlandet, Smart Nest, Suradnici u učenju, Sykli, Universitat Autonoma de Barcelona, Universität Graz, Vilnius Ozas Gymnasium
Translator(s)	Evelyne Putz, Julian Priebsch, Universität Graz
Graphic design	Cristina Obae & Merja Salminen, Sykli; Lidija Kralj, Suradnici u učenju; Evelyne Putz, Universität Graz
Image credits	The cover page image was AI generated using Canva. It complies to Canva Acceptable Use Policy and Canva's Content License Agreement.

The EducationalAI project is funded by the European Union. Neither the European Union nor the European Education and Culture Executive Agency (EACEA) can be held responsible for the content of this publication

More information about the Erasmus+ programme: www.oph.fi/erasmusplus

Table of Contents

Introduction 5
Target Audience and Objectives 5
Specific Features of the Education Sector 6
What is Artificial Intelligence (AI)? 6
Ethics and Law 7
Part I: Ethical Aspects 9
Introduction 9
What is AI Ethics? 9
Underlying Principle: The Human-Centred Approach 9
Ethical Responsibility Throughout the Entire Lifecycle 10
Taking the Context into Account: Education 10
Tensions and the Need for Balance 11
Ethical Principles 11
Respect for Human Autonomy 11
Safety and Harm Prevention 13
Fairness, Inclusion, Equal Opportunities, and Non-Discrimination 17
Human Oversight 20
Explainability and Transparency 22
Privacy and Data Protection 24
Accountability and Complaint Mechanisms 26
Staff Involvement and Participation 28
Respect for the Needs of Young People and the Social Role of Educational Institutions 29
Academic Integrity 31
Illustration: Ethical Considerations at a Glance 34
Part II: Legal Aspects 35

Introduction 35
The AI Act 36
Introduction 36
Which Systems are Covered by the AI Act? 37
Key Roles of Educational Institutions Under the AI Act 41
AI Literacy as a Fundamental Requirement 44
The Four Risk Categories and Their Requirements 47
General-Purpose AI Models 57
Data Protection Requirements 60
Introduction 60
Legal Framework 61
Common Practical Challenges 65
Recommendations for Data Protection-Compliant Use of AI 66
Children's Rights and the Use of AI in Education 69
Illustration: Key Legal Aspects at a Glance 71
Conclusion 72
Annex: The Guide in a Nutshell 73
Ethical AI Integration in Education 73
Essential Aspects 73
Step 1: Before Deployment 74
Step 2: While Deploying 75
Step 3: After Deployment 76
Lawful AI Integration in Education 77
Understanding the Legal Context of AI Use 77
Step 1: Compliance with EU Law 77
Step 2: Compliance with National and Regional Legal Requirements 80
Step 3: Compliance with Requirements of Public Authorities 80

References 81

Introduction

Target Audience and Objectives

This guide is aimed at **educational leaders** who are already using artificial intelligence (AI) or are planning to do so – whether in teaching or in administrative processes. In this context, "educational leaders" refers to both educators (e.g., teachers or university professors) and management personnel (e.g., school principals or rectors) who shape how AI is used in education.

The guide addresses key ethical and legal questions that arise (or may arise) when using AI in the education sector. Its aim is to offer guidance in navigating these challenges and to provide practical advice for the responsible, reflective, and legally compliant use of AI, with a focus on educational leaders.

The first part of the guide introduces important **ethical principles** that must be taken into account when implementing AI in education. The second part focuses on **selected legal frameworks** that are particularly relevant in this context: the EU Artificial Intelligence Act (AI Act, as of 2025), the General Data Protection Regulation (GDPR), and the role of children's rights (notably Article 24 of the Charter of Fundamental Rights of the EU, CFR).

While the guide focuses on central ethical and legal aspects, it cannot cover all areas related to the use of AI in educational settings. Topics such as intellectual property, labour law, or the legal consequences of academic dishonesty, for instance, are not addressed.

Key definitions, important notes, and examples appear in coloured boxes. At the end of each chapter, guiding questions support readers in reflecting on the content and applying it in practice. The annex, *The Guide in a Nutshell*, offers a concise overview of the guide's main points, underscoring the most important considerations.

Specific Features of the Education Sector

This guide takes into account the specific conditions of the school and broader education sector, which involve both adults and young people aged 14 and above. Depending on the target group, needs may vary:

- Children and adolescents are, at least in part, more vulnerable and in need of protection than adults.
- Relationships of authority exist between educators and learners as well as between leaders and staff.
- Information asymmetries frequently exist between the stakeholders involved.

What is Artificial Intelligence (AI)?

The term "artificial intelligence" (AI) has been in use since the 1950s[1] and continues to be defined differently across various contexts. This guide follows the **definition provided in the AI Act**,[2] which establishes a uniform legal framework "*in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems) in the Union*".[3] As a regulation, it is directly applicable and binding in all Member States.

[1] For a historical background, see R Kline, 'Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence' (2011) 33(4) *IEEE Annals of the History of Computing* 5-16 <https://ieeexplore.ieee.org/document/5477410> (last accessed 26 August 2025).
[2] Regulation (EU) 2024/1689 of the European Parliament and the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) OJ L 2024/1689. The full text is available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (last accessed 26 August 2025).
[3] Recital 1 AI Act.

Definition: An AI system as defined in Article 3(1) of the AI Act is "*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*".

This definition is very broad: it covers a wide range of technologies, concepts, and models that, depending on their technical design, may operate using methods such as machine learning or logic- and knowledge-based approaches.[4] This enables a wide spectrum of applications, such as:

- image or text generators
- translation systems
- facial or emotion recognition systems
- recommendation systems, and
- administrative tools with AI components

The diversity of possible applications – and the associated opportunities and risks, particularly in the educational context – makes it essential for educational leaders to engage with the ethical and legal foundations of AI use.

Ethics and Law

Legal requirements for the use of AI overlap in some areas with ethical principles, but they must be clearly distinguished. Adhering to ethical guidelines, such as those set out in the first

[4] Recital 12 AI Act; European Commission, 'Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)' C(2025) 5053 final, available at: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> (last accessed 26 August 2025).

part of this guide, does not automatically ensure legal compliance. A careful review of the legal framework is therefore always necessary.

Part I: Ethical Aspects

Introduction

What is AI Ethics?

AI ethics is a branch of applied ethics concerned with questions and decisions related to the design, development, and use of AI systems.[5] Central themes include respect for human autonomy, fairness, transparency, and accountability.

Underlying Principle: The Human-Centred Approach

At the heart of AI ethics is a human-centred approach: the use of AI systems should not be guided primarily by economic interests or technical feasibility but by the **well-being and empowerment of the individuals affected**.[6] In the educational context, these individuals include, on the one hand, learners and (where applicable) their parents or guardians, and on the other hand those working in teaching and administration. Their needs and goals – as well as the broader societal importance of education – should take priority. When applied ethically,

[5] See, for instance, C Müller, 'Ethics of Artificial Intelligence and Robotics', in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Summer edition, 2025) <https://plato.stanford.edu/entries/ethics-ai/> (last accessed 26 August 2025).

[6] See, for example, High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (April 2019) 4, 37, hereinafter cited as HEG AI, 'Ethics Guidelines', <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (last accessed 26 August 2025); UNESCO, 'Guidance for generative AI in education and research' (2023) 18, 29, 38, hereinafter cited as UNESCO, 'Guidance for GenAI', <https://unesdoc.unesco.org/ark:/48223/pf0000386693> (last accessed 26 August 2025); Deutscher Ethikrat [German Ethics Council], 'Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme' [Humans and Machines – Challenges Posed by Artificial Intelligence, Opinion] (2023), hereinafter cited as Deutscher Ethikrat, 'Mensch und Maschine', 38-39, 246 <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf> (last accessed 26 August 2025). See also *Education Recommendation 1* of the German Ethics Council in Deutscher Ethikrat, 'Mensch und Maschine' 41.

AI in education can promote inclusive and high-quality learning and support lifelong education.

Ethical Responsibility Throughout the Entire Lifecycle

Ethical principles should be considered from the outset and throughout the entire lifecycle of an AI system: from its initial design and technical development to its practical implementation. [7] This can be achieved through various measures, such as **technical safeguards** or **organisational approaches** like guidelines and certifications. "By-design" concepts – such as "privacy by design" or "security by design" – integrate key ethical and legal considerations already at the design stage. Many AI systems can evolve continuously during their use (e.g., through machine learning). This requires **ongoing attention**: Regular reviews may be needed to ensure that systems continue to meet ethical standards. The rapid pace of technological advancement in the AI field further reinforces the importance of regularly revisiting and, if necessary, revising earlier decisions about system use.

For educational institutions, **technical criteria** are especially important when selecting AI systems. For daily, practical use, however, organisational frameworks such as **guidelines** play a more central role.

Taking the Context into Account: Education

AI technologies and their potential applications in education are highly diverse. Which ethical considerations are most relevant depends both on **the AI system's capabilities and**

[7] See, for instance, UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' (2022), hereinafter cited as: UNESCO, 'Recommendation on the Ethics of AI', <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (last accessed 26 August 2025).

functionality (e.g. whether it is generative or not), but also on the **specific use case** (e.g., creating teaching materials or direct interaction of an AI system with learners).

The **target group** is also crucial: In the case of learners, factors such as age, socioeconomic background, gender, and any learning difficulties or disabilities play a significant role.

> **Example:** An outline for an essay is to be created with the help of a generative AI system. The ethical considerations will differ depending on whether the learners involved are around 12 or 17 years old.

For teaching and administrative staff in educational institutions, the use of AI systems may impact **professional roles**, **areas of responsibility**, and **working conditions**.

Tensions and the Need for Balance

In practice, ethical principles may come into conflict with one another. In such situations, **careful balancing** is required.

> **Example:** The use of an AI-based tool to assist in monitoring learners during an exam might improve fairness, but it could also significantly infringe on their privacy.

Ethical Principles

Respect for Human Autonomy

The individual freedom to make self-determined decisions, to take responsibility for them, and to act accordingly must be upheld. **AI systems should strengthen this freedom, not**

undermine it.[8] This applies equally to learners, parents and guardians, educators, and those working in administration or management.

Key Points:

- **Choice and agency:** Affected individuals – including children and adolescents – should be able to preserve their autonomy. Where possible and meaningful, they should be able to influence decisions and choose alternatives.
- **No manipulation:** AI systems must not engage in covert control or emotional influence.
- **Promoting AI literacy:** This involves strengthening the ability of individuals to critically assess the technologies in use, understand their impact, and engage with them responsibly. For children, adolescents, and other vulnerable groups, **age-appropriate and developmentally appropriate support and information** are essential.

Example: A school introduces an AI-based system for grading exam papers. The system automatically suggests specific grades, which are then applied without further review. Educators may only deviate from these suggestions in exceptional cases, and learners have no option to request a review of the AI-generated grade.

In this case, neither learners nor educators are granted any decision-making space, and their autonomy is restricted: Learners cannot question the assessment, and educators lose the ability to make pedagogical decisions regarding grading.

If AI systems are used for performance assessment, this must always be done with reflection and care. The pedagogical experience and professional knowledge of educators form the foundation of fair evaluation. Learners must have the opportunity to contribute their perspective and question the results.

Note on this example: Since this system is used as intended in the education sector for assessing learning outcomes, it could qualify as a high-risk AI system under the AI Act. More information can be found in this guide's chapter on the AI Act.

[8] See, for instance, HEG AI, 'Ethics Guidelines' 12, 16.

Important: Legal provisions may also apply in this context — such as requirements regarding parental or guardian consent!

Guiding Questions:

- Can all affected stakeholders (e.g., learners, educators, administrative staff) participate in shaping the use of the AI system or choose alternatives?
- Is individual decision-making of all stakeholders appropriately taken into account — e.g., by offering options or alternative approaches?
- Do affected stakeholders receive sufficient (including age-appropriate and developmentally appropriate) information on how the AI system works, what it means, and what its effects may be?
- Is there a risk that individuals place too much trust in the AI system (e.g., uncritically following AI-generated suggestions) or become dependent on it?

Connections:

The principle of respect for human autonomy...

- ➔ is closely linked to other ethical principles – particularly *explainability and transparency*, *accountability and complaint mechanisms*, *staff involvement and participation*, and *respect for the needs of young people and the social role of educational institutions*.
- ➔ is also supported by the *right of children to be heard and included*.
- ➔ is reflected in various legal provisions, such as the *obligation to build AI literacy* under Article 4 of the AI Act, as well as in the *transparency requirements* under Article 50 of the AI Act.

Safety and Harm Prevention

The principle of "*do no harm*" is a fundamental ethical guideline. **Harms caused by AI systems must be prevented or minimised.**[9]

[9] See, for instance, also in HEG AI, 'Ethics Guidelines' 12.

Potential risks include:

- Misinformation or misleading content, psychological stress, or impairments in social development and skills.
- Manipulative effects: Interactive systems, such as large language models, can simulate human-like conversation, empathy or emotional connection. This creates a risk of emotional manipulation, particularly if users interpret the AI system's responses as genuine or personal.
- Lack of technical security: Systems must be technically robust (preventing unintended errors) and secure against targeted attacks (e.g., through hacking, misuse, or abuse).
- Unreliable outputs: Some AI systems may produce content that sounds plausible, but is actually incorrect or misleading (commonly referred to as "hallucinations").[10] This can, for example, reduce quality and reliability of educational materials or decisions.

A careful risk assessment must be carried out before any AI system is deployed. This is particularly important in educational settings such as schools, where trust and protection play a central role.

For the preliminary assessment and for decision on appropriate safety measures, the following factors must be considered in particular:

- **functionality and capabilities** of the AI system
- **context** and specific **use case** (e.g., teaching, administration), considering:
    - potential power or information imbalances (e.g., between educators and learners)
    - characteristics of the affected persons or groups (e.g., vulnerable groups such as children and adolescents)
- type of **data** involved (e.g., sensitive or personal data)
- **possible negative impacts** – especially on children and adolescents – such as:
    - on mental wellbeing
    - on natural behaviour
    - and on social development

[10] See, for instance, G Perković, A Drobnjak and I Botički, 'Hallucinations in LLMs: Understanding and Addressing Challenges' *2024 47th MIPRO ICT and Electronics Convention (MIPRO)* (2024) 2084–2088 <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10569238> (last accessed 26 August 2025).

In the event that problems occur, **response measures** must be in place to prevent or mitigate harm.

Another aspect of harm prevention is the consideration of **sustainability and environmental friendliness** in the development, deployment, operation, and use of AI systems.[11] Relevant here are, for example, the resource and energy consumption of AI systems, particularly during training.

Example: An educator instructs learners to use a generative AI system to develop ideas for creative writing tasks. However, no prior check is carried out to ensure the AI system is restricted to age-appropriate content and pedagogically suitable. In some cases, the AI system generates content involving violence.

In this example case, the AI system was applied without adequate reflection. Its suitability for the intended target group was not assessed, meaning the principle of harm prevention was not sufficiently taken into account. As a result, there is a risk that problematic outputs from the system could negatively affect learners – especially in terms of their psychological well-being and social development.

AI tools must be tested before being used in the classroom, and a risk assessment must be carried out — particularly concerning age-appropriate and developmentally suitable content.

Important: Legal regulations concerning safety measures – particularly regarding the technical design and intervention options – must be strictly observed when using AI systems.

Guiding Questions:

- Risk assessment:
  - Is there a clearly designated individual or team responsible for assessing potential risks prior to AI use?

[11] See also in HEG AI, 'Ethics Guidelines'.

    - Has this assessment been carried out and documented? Were social impacts and specific vulnerabilities (e.g., age, emotional development) taken into account?
- Have those using AI received the necessary training?
- Guidelines:
    - Are there clear guidelines for those using the AI system (e.g., educators)?
    - Do the guidelines, for example, specify:
        - Who is authorised to use AI?
        - Which tools are approved for use?
        - For which purposes they may be used (e.g., for research, inspiration, analysis of certain data sets)?
        - What types of data may be entered into the AI tool?
        - Whether users must complete (additional) training before they are permitted to use AI?
        - Who is responsible for reviewing AI outputs for aspects such as accuracy, potential bias, alignment with ethical standards?
        - Whether and how AI use must be documented?
        - Whether and how AI-generated content must be labelled?
    - Do the guidelines address potential risks?
    - Are the guidelines regularly reviewed and updated?
- Are there technical and organisational safeguards for safety and error prevention?
- Have the system provider's safety protocols been checked and are they being followed?
- Are there defined procedures to respond promptly to problems or misuse?

Connections:

The principle of safety and harm prevention...

➔ is closely linked to other ethical principles – particularly *fairness, inclusion, equal opportunities, and non-discrimination*, *human oversight*, *explainability and transparency*, *privacy and data protection*, *accountability and complaint mechanisms*, *staff involvement and participation* as well as *respect for the needs of young people and the social role of educational institutions*.

- is embedded in legal obligations, such as the differentiated responsibilities for providers and deployers of AI systems depending on *risk classification* under the AI Act.

Fairness, Inclusion, Equal Opportunities, and Non-Discrimination

All people should be treated equally and fairly and have access to the same opportunities. AI systems must be designed and used in ways that **promote fairness, respect diversity, enable inclusion, and avoid reproducing or reinforcing any form of discrimination** – including by ensuring algorithmic fairness.[12] Particular focus must be placed on:

- **Equal access and non-discriminatory use:** Systems must be inclusive and accessible, regardless of characteristics such as gender, ethnicity, cultural or socioeconomic background, or age.
- **Avoiding bias:** In practice, bias frequently occurs. This refers to systematic prejudice, for example, against a person or group. Bias can originate from algorithmic design, from distortions present in training data, or from data collected during use of the AI system. Therefore, already during system development and selection, care must be taken to ensure that training is based on diverse and representative data sets.
- **Complaint mechanisms:** These allow users to provide feedback and correct issues.

[12] See, for example, HEG AI, 'Ethics Guidelines' 12-13, 18-19; UNESCO, 'Guidance for GenAI' 24; UNESCO, 'Recommendation on the Ethics of AI' 20-21.

Example: A school implements an AI-based learning system intended for use both in class and at home. Learners need their own technical equipment such as personal digital devices to use the system. Some learners cannot afford these due to high costs, and no loan devices are available. As a result, these learners cannot access the learning system and are excluded from important content.

This is an example of unequal access. Socioeconomic disparities were not taken into account, and some learners are disadvantaged and excluded from the learning process. In contrast, fairness and equal opportunities in education require that all learners have equitable access to learning tools, and that differences in starting conditions are addressed. Possible solutions in such cases may include providing loan devices, offering alternative access options (e.g., offline versions of the learning system), or implementing pedagogical alternatives.

Used responsibly, AI can also **enhance access to education** and promote the inclusion of persons with disabilities or special needs.[13]

Examples:

- A university uses an AI system that generates real-time subtitles for lectures. Learners with hearing impairments can follow the content without delay.
- An AI system detects images in presentations and immediately provides an audio description. This enables learners with visual impairments to access diagrams or photographs directly.

Such uses of AI can support or enable participation in educational activities.

Guiding Questions:

- Access:
  - Do all users have equal access to AI-supported services?

[13] See also *Education Recommendation 2* of the German Ethics Council in Deutscher Ethikrat, 'Mensch und Maschine' 41.

  - Have potential barriers been identified (e.g., lack of access to devices due to socioeconomic background, language barriers) and have solutions been implemented?
  - Is the AI system itself designed to be accessible and inclusive for people with disabilities or special needs? If not: Can existing barriers be removed?
- Is it ensured that the AI system does not contain or reproduce biases?
- Is the system applied fairly to all individuals (e.g., not disproportionately to already disadvantaged groups)?
- Complaint mechanisms:
  - Can individuals affected by perceived inequality file complaints or give feedback?
  - Are there clear procedures and responsibilities for processing complaints and feedback?
  - Are affected individuals sufficiently informed about these options?

Connections:

The principle of fairness, inclusion, equal opportunities, and non-discrimination...

- ➔ is closely linked to other ethical principles – particularly *safety and harm prevention*, *human oversight*, *explainability and transparency*, *accountability and complaint mechanisms*, as well as *respect for the needs of young people and the social role of educational institutions*.
- ➔ is also reflected in legal requirements, such as the requirement that *training, validation, and testing data sets for high-risk AI systems must be representative*, as set out in Article 10(3) AI Act.

Human Oversight

Essential decisions must not be left entirely to automated processes. **Oversight must always remain with humans** who monitor, influence, correct, or interrupt the operation and outputs of AI systems when necessary.[14]

> **Example:** An AI system is used to assess the suitability of applicants for continuing education programmes. It analyses professional experience and educational background. The results automatically lead to some individuals receiving access to specific training pathways, while others are excluded. There is no review by human professionals, and the admission decision is final.
>
> In this case, the AI system makes a decisive determination about the educational opportunities available to applicants. There is no human oversight or option for correction. This violates the principle of human oversight by delegating a crucial decision to an automated process. Admission decisions for educational programmes should instead be made by human professionals. AI systems can be used to support these decisions, but the final responsibility must rest with a person.
>
> **Note on this example:** The use of AI systems in the context of access, admission, or assignment to educational opportunities may be classified as high-risk under the AI Act. In such cases, stricter requirements apply! More information can be found in this guide's chapter on high-risk AI systems under the AI Act.

Depending on the system and use case, human oversight can take different forms. Three models are usually distinguished:[15]

- **Human-In-The-Loop (HITL):** Humans are actively involved in the decision-making processes of the AI system — e.g., through manual approval, review, or correction of results. They can intervene, adjust outcomes, or stop the process.

[14] Compare, for example, HEG AI, 'Ethics Guidelines' 16; UNESCO, 'Recommendation on the Ethics of AI' 22.
[15] Compare, for example, HEG AI, 'Ethics Guidelines' 16; L Methnani, A A Tubella, V Dignum and A Theodorou, 'Let Me Take Over: Variable Autonomy for Meaningful Human Control' (2021) 4 *Frontiers in Artificial Intelligence* Article 737072 <https://doi.org/10.3389/frai.2021.737072> (last accessed 26 August 2025).

- **Human-On-The-Loop (HOTL):** The AI system operates fundamentally independently but is continuously monitored. If anomalies or malfunctions arise, the processes can be intervened in, results can be changed, or the process can be stopped.
- **Human-In-Command (HIC):** A human retains overall control of the operation of the AI system. They decide whether and how the system is deployed, and have the authority to shut it down, and overturn individual results.

The appropriate form of oversight **depends on the specific use case and associated risks**. The more sensitive the use case and the greater the potential risk, the more robust human oversight should be.

Guiding Questions:

- What risks are associated with the use of the AI system?
- Is the use case sensitive — e.g., performance evaluations or decisions about educational pathways?
- Which oversight model is appropriate in light of these risks?
- At which stages is human oversight required — e.g., at the start, at intervals, continuously?
- Responsibilities and processes:
    - Is it clearly defined who is responsible for oversight and for initial and ongoing risk assessments? What rules and procedures apply?
    - Does this person have the necessary expertise to exercise this role responsibly, or is additional training required?
    - Can affected individuals (e.g., educators or learners) raise concerns, and are they informed about this option?
    - Are there clear procedures to appropriately respond to problematic incidents?

Connections:

The principle of human oversight...

- ➔ is closely linked to other ethical principles – particularly *safety and harm prevention*, *fairness, inclusion, equal opportunities, and non-discrimination*, *explainability and transparency*, and *accountability and complaint mechanisms*.
- ➔ is reinforced by the legal framework governing the use of AI, such as *specific requirements for ensuring human oversight in high-risk AI systems* under Article 14 AI Act.

Explainability and Transparency

Explainability and transparency in the context of AI systems enable affected individuals to make informed decisions, properly assess risks, and exercise their rights.[16]

Definition: Explainability means that the technical processes, capabilities, limitations, and objectives of an AI system **can be described in a way that is understandable**.

Not all AI systems are easily explainable. Systems based on machine learning are sometimes considered "black boxes": The precise decision-making pathways, model structures, and parameters are often technically difficult to understand or reconstruct. In education, it is particularly important to carefully assess whether the use of such systems is appropriate.

Definition: Transparency means that **affected persons are informed about the use and functioning of an AI system**. It specifically includes:

- That the AI system is **identifiable** as such (e.g., in the case of chatbots),
- Information about the system's **functions**,
- **When, how, and for what purpose** the system is used,
- What **significance** the outputs have (e.g., how they influence assessments or recommendations).

Transparency also means **tailoring the information** to the **specific use context** and the **affected persons**: Are they children, adolescents, adults, or individuals with disabilities or special needs?

Important: Legal obligations, such as those regarding data protection and the confidentiality of security-sensitive, technical or company-specific information, must always be observed. Under the AI Act, specific transparency obligations may apply; see this guide's chapter on the AI Act for further details.

[16] Compare, for example, HEG AI, 'Ethics Guidelines' 18; UNESCO, 'Guidance for GenAI' 22; UNESCO, 'Recommendation on the Ethics of AI' 22.

Example: A school administration uses an AI system to predict learners' likelihood of success for future educational pathways using information about the learners' grades and absences. These predictions feed into internal recommendations for allocation to support measures or educational counselling. Neither the learners themselves nor their parents or guardians are informed that an AI system is being used or which data influence the predictions.

This case demonstrates a lack of transparency: the individuals concerned are unaware that an AI system is involved in important decisions about their educational future. As a result, they cannot see how the system works, what criteria are relevant, or what consequences the predictions may have. The use of the AI system, the data involved, and the key factors in decision-making must be disclosed.

Note on this case example: This use of AI may affect learners' educational trajectories and could – as in the previous example – fall under the high-risk category as defined by the AI Act. Accordingly, stricter requirements may apply.

Guiding Questions:

- Are the basic methods, functions, and decision-making mechanisms of the AI system known? Are they (at least in simplified form) explainable to affected persons?
- Is it clear to those affected when they are interacting with an AI system (e.g., with chatbots)?
- Are affected persons (e.g., learners or educators) informed that an AI system is being used and in what context? Are they told what role the AI system plays in specific decisions?
- Are the provided explanations clear, age-appropriate, and tailored to the context?

Connections:

The principle of explainability and transparency...

➔ is closely linked to other ethical principles – particularly *respect for human autonomy*, *safety and harm prevention*, *fairness, inclusion, equal opportunities, and non-discrimination*, *human oversight*, *privacy and data protection*, *accountability and complaint mechanisms*, and *staff involvement and participation*.

- may at times be in tension with other ethical obligations – such as *data protection* or *security* – for example, when disclosing personal data or security-sensitive details would be inappropriate. In such cases, a careful balancing of interests is needed.
- is reflected in legal obligations, such as the *transparency requirements* under Article 50 of the AI Act.

Privacy and Data Protection

AI systems often **process personal data** – whether from teaching or administrative staff, learners, or their parents or guardians. These systems may also infer additional information based on usage data. Personal data must be handled responsibly.[17] Particular care is required for **sensitive data**, such as information on political opinions, origin, religious or political beliefs, or health status.

An ethically responsible approach to data includes the following core principles:

- **Data minimisation:** Only collect data necessary for the specific purpose.
- **Purpose limitation:** Use the data solely for the purpose for which it was collected.
- **Access restriction:** Only explicitly authorised individuals should have access to the data — and only to the extent necessary for their duties.
- **Security:** Data must be stored securely, both technically and organisationally.
- **Transparency:** Affected persons must be informed about which data are collected, for what purpose, how long they are stored, and how they are used during that time.

AI systems can significantly intrude on privacy, particularly due to their analytical capabilities. Especially critical are systems for facial recognition, behaviour or emotion monitoring, or personality profiling.

Potential threats to privacy and personal data from an AI system **must be evaluated before its use**. Key factors for the preliminary assessment and determining the necessary security measures are detailed in the chapter on *safety and harm prevention*. These are particularly relevant in relation to data protection and privacy risks.

[17] Compare, for example, HEG AI, 'Ethics Guidelines' 17; UNESCO, 'Recommendation on the Ethics of AI' 21-22.

Important: The protection of privacy and personal data is governed by binding legal frameworks – most notably the General Data Protection Regulation (GDPR), but also relevant national laws. These regulations may exceed or further specify the ethical principles outlined here. Legal requirements (e.g., on information obligations and consent) must always be respected.

Example: A school introduces an AI system to support teachers. The system uses cameras installed in classrooms to analyse learners' facial expressions, gestures, and gaze direction. It draws conclusions about attention levels and shares this information with educators so they can, for example, adjust their teaching style.

This application significantly intrudes upon learners' privacy. It may also have other negative impacts (e.g., on mental well-being, natural behaviour, and social development). For ethical reasons alone, such a system should not be used. Instead of automated surveillance and analysis, educators should rely on their own observations, proven pedagogical methods, and trust-based relationships, enabling open dialogue around attention and engagement.

Note on this case example: This application is not only ethically inappropriate but also legally prohibited. The AI Act prohibits the use of AI systems "to infer emotions of a natural person in the areas of workplace and education institutions". Further information on prohibited AI practices can be found in the relevant chapter of this guide.

Guiding Questions:

- What data does the AI system collect or process? Does this include personal or particularly sensitive data?
- Can the same goal be achieved using less or less specific data?
- Is it possible to anonymise or pseudonymise the data?
- Are those affected clearly informed about the type, purpose, duration, and use of their data?

- Are the data used exclusively for the purpose for which they were collected?
- Who has access to the data? Are they authorised and trained, and do they need access for their responsibilities?
- Are the data deleted when they are no longer needed?

Connections:

The principle of privacy and data protection...

- ➔ is closely linked to other ethical principles – particularly *safety and harm prevention*, *explainability and transparency*, and *accountability and complaint mechanisms*.
- ➔ is reflected in numerous legal provisions, such as *the GDPR requirements for handling personal data* and the AI Act's *rules on prohibited AI practices and high-risk AI systems*.

Accountability and Complaint Mechanisms

Even when technology is used to support or automate processes, **human responsibility for the design, selection, and application of AI systems must be maintained**.[18]

Key aspects include:

- **Clear allocation of responsibilities** – who is responsible for selecting, assessing, implementing, monitoring and operating the AI system.
- **Traceability:** Decisions made with the help of or based on AI must be explainable, verifiable, and correctable if necessary.
- **Provision of guidelines** on the use of AI systems by the organisational leadership or community (e.g., educators, learners, parents/guardians).
- **Competence of responsible persons** (e.g., through appropriate continuing education).
- **Mechanisms for feedback and complaints** when the AI system works incorrectly, delivers inexplicable or unfair results, or inappropriately intrudes on privacy.

[18] See also, for example, HEG AI, 'Ethics Guidelines' 19-20.

Example: A teacher uses an AI system to grade learners' essays. The system assigns scores automatically, and the teacher adopts them without review. When questioned by learners and parents, the teacher refers to the system. Following complaints, the school leadership defends the approach as pedagogical discretion.

In this case, responsibility is neither accepted by the teacher nor assigned by the school leadership — instead, it is effectively shifted to the AI system. This creates a gap in accountability. Learners cannot question the results or identify a responsible human decision-maker. This fails to meet ethical standards: roles and responsibilities in grading must be clearly defined, AI systems should support but never replace human judgment.

Guiding Questions:

- Are there clearly named individuals within the institution responsible for selecting, implementing, and monitoring AI systems? Do they have sufficient expertise in ethical, pedagogical, and legal matters?
- Are there clear rules on how much influence AI-generated results can have on human decisions (e.g., performance assessments)?
- Do all involved understand that they remain responsible for their decisions – even when supported by AI?
- Are AI-influenced decision-making processes documented? Can decisions later be traced and justified, including identifying who made them?
- Do affected persons have a way to raise concerns? Are they informed about this option and the relevant contact point?
- Are there clear procedures for processing complaints and feedback?

Connections:

The principle of accountability and complaint mechanisms...

- ➔ is closely linked to other ethical principles – particularly *respect for human autonomy*, *safety and harm prevention*, *fairness, inclusion, equal opportunities, and non-discrimination*, *human oversight*, *explainability and transparency*, *privacy and data protection*, *staff involvement and participation*, and *respect for the needs of young people and the social role of educational institutions*.
- ➔ is also supported by legal provisions, such as Article 22 GDPR (*protection against decisions based solely on automated processing*), and Article 14 AI Act (*requirements for ensuring human oversight in high-risk AI systems*).

Staff Involvement and Participation

Ethically sound AI deployment requires the **early and continuous involvement of all affected personnel – educators, administrative staff, and those in managerial positions –** in decisions on the introduction and use of AI technologies.[19] This inclusive approach helps ensure that different perspectives are considered and that **working conditions at all levels are not negatively affected**. Educators, administrative staff, and others should:

- Be **informed and actively involved**,
- Have **access to training** (covering technical basics, e.g., functionality, limitations and risks, but also legal and ethical aspects),
- Be able to **raise concerns and make suggestions** for improvement.

Their assessments of risks, benefits and impacts should be **appropriately considered**.

The use of AI systems should:

- **not replace** educators or other staff, but support, complement, and relieve them in their work;[20]
- **not impair** social interaction between learners and educators, which is essential for understanding learners' needs and progress.

Example: A school leadership team implements an AI system to automatically generate timetables and assign teachers. While it takes resource constraints into account, it ignores established team structures, teachers' pedagogical approaches, and their individual obligations. No consultation with the teachers takes place.

Here, AI is introduced with a focus on efficiency but without involving those affected. Social and pedagogical factors are neglected. In education, such systems should be implemented collaboratively, respecting educators' perspectives and professional judgment.

Guiding Questions:

- Are staff members informed early about planned AI deployments?

[19] See, for instance, HEG AI, 'Ethics Guidelines' 19.
[20] See also *Education Recommendation 10* of the German Ethics Council in Deutscher Ethikrat, 'Mensch und Maschine' 44.

- Do they have opportunities to give feedback, especially on ethical or pedagogical concerns?
- Do staff at all levels receive understandable information about how the system works, including its limitations and risks? Are appropriate training opportunities provided?
- Is the selection and implementation of AI systems designed to recognize and strengthen educators' and other staff's professional roles?
- Is there a risk that important insights into learners' needs and progress could be lost due to AI use?

Connections:

The principle of staff involvement and participation...

- ➔ is closely linked to other ethical principles – particularly *respect for human autonomy*, *safety and harm prevention*, *explainability and transparency*, *accountability and complaint mechanisms*, as well as *respect for the needs of young people and the social role of educational institutions*.
- ➔ is also supported by legal provisions for AI use: in particular, the AI Act provides for *information of workers before the use of high-risk AI systems at the workplace*.

Respect for the Needs of Young People and the Social Role of Educational Institutions

The use of AI systems in education often affects children, adolescents and young adults – **individuals at formative stages of their lives**. The impact of AI on the development of social and cognitive abilities of children and young people of different age groups must therefore be carefully assessed.

Educational institutions serve an **essential social function**: interpersonal interaction fosters not only learning but also social, emotional, and personal development. Collaboration among learners and between learners and educators is vital for cultivating important skills such as critical thinking, empathy, teamwork, and constructive conflict resolution. Technology must support – not undermine – this **social dimension**.[21]

[21] See, for example, Deutscher Ethikrat, 'Mensch und Maschine' 40, 221-224.

AI systems used with children, adolescents and young adults must therefore:

- be **age-appropriate** in design,
- follow a **clear pedagogical concept**,
- and **support learners' cognitive, emotional, and social development**.

Example: A school uses an AI system to group learners based on performance and task completion speed to optimize group efficiency. The previous groupings made by a teacher – intentionally mixing stronger and weaker learners – are abandoned. The new groups are homogeneous in learning progress and prior knowledge. However, the lack of mixing reduces peer learning and the assumption of social responsibility for each other.

The AI-supported group allocation does not consider pedagogical objectives such as mutual support among learners and social mixing, but only efficiency considerations. This can hinder important social learning processes and fix pupils in their social roles. Instead of automated allocations, groupings should be guided by educational objectives and teachers' experience. Social aspects should be considered, and learners themselves should receive appropriate opportunities for participation.

Guiding Questions:

- Has the AI system been tested and found suitable for use in education and for the relevant age group?
- Is it ensured that contents are age-appropriate, particularly in interactive AI systems?
- In cases of direct interaction with learners: Does the AI system communicate clearly and age-appropriately that any "social" behaviour is simulated and that it has no human emotions?
- Does the AI system reflect pedagogical goals and support the development of skills such as critical thinking and social interaction?
- Is the AI system designed to strengthen – not undermine – the pedagogical role of educators?
- Is there a risk that learners might rely too heavily on the system (e.g., in discussions or for problem-solving)?

- Does the use of the AI system limit interaction between learners or between learners and educators?
- Is the social role of the educational institution as a developmental space for young people affected by the system's use?

Connections:

The principle of respect for the needs of young people and the social role of educational institutions...

➔ is closely linked to other ethical principles – particularly *respect for human autonomy*, *safety and harm prevention*, *fairness, inclusion, equal opportunities, and non-discrimination*, *accountability and complaint mechanisms* and *staff involvement and participation*.

Academic Integrity

The growing availability of AI tools – and their ability to generate content like texts, images, or solutions to mathematical problems – has introduced new challenges for educational institutions. One of these is how to uphold principles of honesty and fairness in learning, often referred to as **academic integrity** or **academic honesty**. This includes not only academic writing but also other tasks for educational purposes, such as homework or exams. A particular concern is the unauthorised use of AI for assignments, projects or exams, with learners presenting the output as their own work. Maintaining academic integrity is essential to ensure that assessments genuinely reflect learners' abilities, knowledge, and skills. However, technological tools for detecting AI-generated content are still in their infancy, and their reliability is widely debated, with experts warning against overreliance on such solutions.[22]

[22] See also European Commission: European Education and Culture Executive Agency, 'AI report – By the European Digital Education Hub's Squad on artificial intelligence in education' (2023) 94 <https://data.europa.eu/doi/10.2797/828281> (last accessed 26 August 2025); Schola Europaea, 'Legal and pedagogical guidelines for the educational use of generative artificial intelligence in the European Schools', hereinafter cited as Schola Europaea, 'Legal and pedagogical guidelines', (Office of the Secretary-General, Pedagogical Development Unit, Ref 2025-01-D-66-en-2, 2025) 17 <https://www.eursc.eu/BasicTexts/2025-01-D-66-en-2.pdf> (last accessed 26 August 2025). See also UNESCO, 'Guidance for GenAI' 28.

Possible Measures:

- **Promoting awareness of academic integrity:** Learners – and, where appropriate, their parents or guardians – should be informed about the importance of academic honesty, including the implications of unauthorised AI use.
- **Setting clear expectations:** Institutions and educators should define and communicate clear rules on when and how learners may use AI tools, and when their use is prohibited. This may include integrating AI guidelines into institutional policies or codes of conduct, and requiring learners to declare whether and how AI tools were used in assignments.[23]
- **Providing clear guidelines and support for educators:** Institutional leadership should establish transparent procedures for investigating and responding to suspected breaches, ensuring fairness for all parties involved.
- **Offering training for educators:** Educational institutions should equip staff with knowledge and resources to recognise AI-generated work and to handle disputes confidently.
- **Considering preventive assessment formats and built-in features:** Some methods such as oral presentations, in-class work, or staged submissions (e.g., outlines and drafts) could reduce opportunities for AI misuse. Some applications also offer built-in version history features which may make it possible to track changes made to the document and to recognise sudden and unusual additions.[24]

[23] See, for instance, the generative AI disclosure template in Schola Europaea, 'Legal and pedagogical guidelines' 18-19.
[24] See also Schola Europaea, 'Legal and pedagogical guidelines' 18.

Example: A school has a clear rule that AI tools may only be used for specific brainstorming tasks when explicitly permitted by the teacher. For an essay assignment, learners were told to complete the work independently, without AI assistance. One submission, however, shows signs of having been largely generated by an AI tool. When confronted, the student denies unauthorised use and challenges the teacher to "prove it". This places the educator in a difficult position.

Possible approaches:

In such cases, it is helpful if the school has established procedures that allow teachers to document their concerns and initiate a standardised investigation process, ensuring transparency and fairness. The institution should also support staff in communicating with learners and parents/guardians, reinforcing the importance of academic honesty.

To help prevent such incidents, the school might also consider alternative assessment approaches – for example, requiring students not only to submit the final essay but also to demonstrate stages of their work during the preparation process (e.g., drafts, outlines, results of in-class work).

Guiding Questions:

- How does the institution ensure that learners understand the importance of academic honesty?
- Are there clear, well-communicated rules on permitted and prohibited AI use for learners, parents/guardians, and staff?
- Does the institution have a transparent process for handling suspected cases of unauthorised AI use?
- Are educators given institutional backing in potential disputes with learners or parents/guardians?
- Have alternative assessment strategies been considered to minimise the risk of AI misuse?

Illustration: Ethical Considerations at a Glance

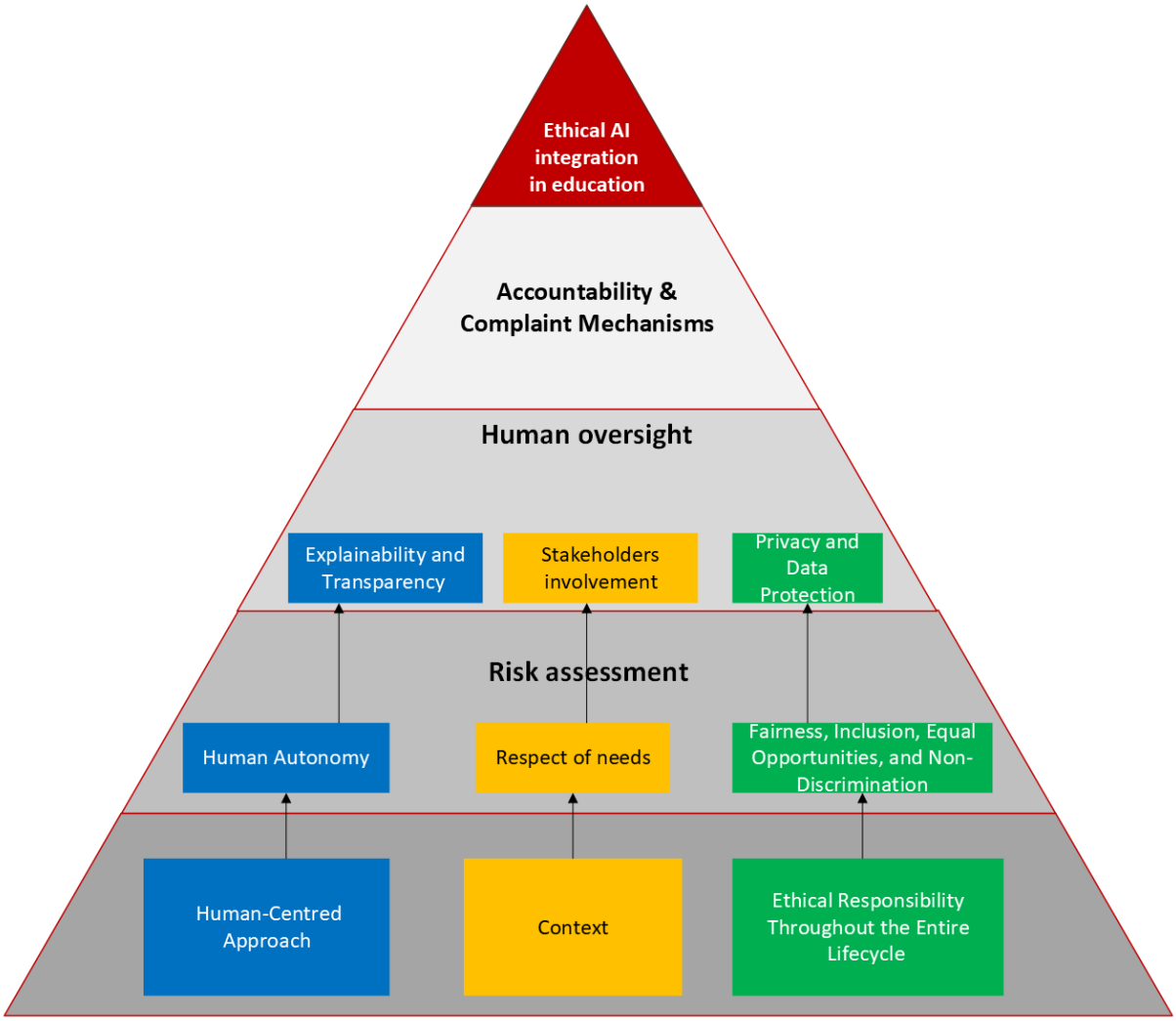

Illustration 1: Ethical AI Integration in Education.

This graph illustrates an approach to ethical AI integration in education. Educational leaders can use it to guide their planning. Viewed from an educator's perspective, the focus will be on curriculum integration; from a management perspective, the emphasis will be on institutional integration of AI.

Part II: Legal Aspects

Introduction

This part of the guide provides an overview of legal provisions particularly relevant for the use of AI in educational contexts by educational leaders, including educators as well as those in management or administrative roles. It covers the following areas:

- regulation under the **EU Artificial Intelligence** Act (AI Act),
- **data protection requirements** (General Data Protection Regulation, GDPR), and
- **children's rights** and AI use in educational contexts.

The aim is to support educators and institutional leaders in the lawful and responsible use of AI systems.

Note: This guide refers to legal requirements at the EU level. However, **national and regional regulations** must also be taken into account – including those issued by the competent educational authorities and, where applicable, by the leadership of the respective educational institution. In addition, both legal provisions and the terms of use of specific AI applications may provide for **age restrictions** that must be observed. The guide does not focus on the legal consequences of academic dishonesty. National legal norms might regulate such behaviour.

The AI Act

Introduction

What is the AI Act?

The **AI Act**[25] is the first comprehensive legal framework for the use of AI in the European Union. As a regulation, it is directly applicable in all Member States of the EU and does not require transposition into national laws.

Objectives and Structure

With the AI Act, the EU pursues several goals: It aims to strengthen the functioning of the internal market while promoting the development and use of human-centric, trustworthy, and safe AI – ensuring a high level of protection whilst simultaneously supporting innovation.[26]

A central element is the **risk-based approach** of the AI Act:[27] Depending on the level of risk posed by an AI system, different legal obligations apply. The regulation is linked to the specific use case of a system. **Four risk categories** are commonly distinguished based on the provisions of the AI Act:

- AI systems with minimal or no risk
- AI systems with limited risk
- High-risk AI systems
- Prohibited AI practices

Each of these categories is explained in the following chapters.

In addition, the regulation contains specific provisions for **general-purpose AI models and systems** – that is, those that were not designed for a specific task, but are suitable for diverse applications.

[25] Regulation (EU) 2024/1689 of the European Parliament and the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) OJ L 2024/1689. The full text is available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (last accessed 26 August 2025).

[26] See, for instance, recitals 1, 2, 3, 8 and Article 1(1) AI Act.

[27] See recital 26 AI Act.

For educational institutions, leaders, and educators, it is essential to understand

- **what role** they assume under the AI Act,
- **which category** AI systems fall into,
- and **which obligations** result from this.

Entry into Force and Implementation

The AI Act entered into force on 1 August 2024. However, its provisions must be implemented gradually.[28] Key dates include:

- Since 2 February 2025, the **prohibition of certain AI practices** and the **obligation to build AI literacy** apply;
- Since 2 August 2025, **provisions on general-purpose AI models** apply;
- From 2 August 2026, **most remaining rules** apply, including those for high-risk AI systems — with the **exception** of those for AI systems embedded in regulated products (these apply only from 2 August 2027).

Which Systems are Covered by the AI Act?

The Seven Elements of an "AI System"

The AI Act applies only to technologies that fall under the definition of an "AI system" as set out in Article 3(1). Support for interpreting this definition is provided by both the **recitals of the regulation**[29] and accompanying **guidelines issued by the European Commission**[30].

[28] Article 113 AI Act. See for a detailed and graphical overview, for example, T Marcelin and L Killmayer, 'AI Act implementation timeline' (At a Glance, PE 772.906, European Parliamentary Research Service, June 2025) <https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/772906/EPRS_ATA(2025)772906_EN.pdf> (last accessed 26 August 2025).

[29] Recitals appear at the beginning of an EU legal act (such as a regulation) and explain the purpose, background, and rationale behind the act. See European Parliament, Council and Commission, 'Joint practical guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation' (2nd edition, 2015) 10, hereinafter cited as: European Parliament *et al*, Joint Practical Guide, <https://op.europa.eu/en/publication-detail/-/publication/3879747d-7a3c-411b-a3a0-55c14e2ba732> (last accessed 26 August 2025).

[30] European Commission, 'Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)' C(2025) 5053 final, available at: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> (last accessed 26 August 2025).

However, these are not legally binding – ultimately, the Court of Justice of the European Union (CJEU) has the final authority on interpretation.[31]

Definition: An AI system as defined in Article 3(1) of the AI Act is "*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*".

This definition names **seven key elements**:[32]

1. An AI system is a **machine-based system**.

- ➔ According to the Commission guidelines, this indicates that AI systems *"are developed with and run on machines"*; the term "machine" includes both hardware elements (e.g., processors, memory) and software elements (e.g., program code, operating systems, algorithms).[33]
- ➔ This means that AI systems *"must be computationally driven and based on machine operations"*.[34] This includes a variety of computing systems – even the most advanced quantum computer systems or biological systems with computing capacity.[35]

2. An AI system is designed to **operate with varying levels of autonomy**.

- ➔ According to the recitals, this means that AI systems "*have some degree of independence of actions from human involvement and of capabilities to operate without human intervention*". [36] An AI system must therefore be able to act independently to a certain extent.
- ➔ This **is the case**, for instance, when a system requires manually provided inputs but can then generate outputs **by itself**, without outputs having to be manually controlled

[31] C(2025) 5053 final para 7; for recitals, see, for instance, European Parliament *et al*, Joint Practical Guide 10.
[32] Siehe C(2025) 5053 final para 9.
[33] C(2025) 5053 final para 11.
[34] C(2025) 5053 final para 12.
[35] C(2025) 5053 final para 13.
[36] Recital 12 AI Act.

or explicitly and exactly specified by a human.[37] In contrast, systems that are "*designed to operate solely with full manual human involvement and intervention*" are **not included**.[38]

3. An AI system **may exhibit adaptiveness** after deployment.

- ➔ As clarified in the recitals, an AI system exhibits adaptiveness after deployment if it possesses self-learning capabilities, i.e., can change while in use and adapt the system's behaviour over time.[39]
- ➔ **Important**: Adaptiveness **is not required** for a technology to qualify as an AI system.[40] It is only a facultative characteristic, not a mandatory requirement.

4. An AI system **infers, from the input it receives, how to generate outputs**.

- ➔ The recitals clarify that this capability goes beyond basic data processing "*by enabling learning, reasoning or modelling*".[41] This capacity for inference distinguishes AI systems from other types of systems.[42] It may be based on various AI technologies — e.g., on machine learning, but also on logic- and knowledge-based approaches.[43]
- ➔ This requirement is **not met**, for example, by systems that "*are based on the rules defined solely by natural persons to automatically execute operations*".[44]

Note: According to the European Commission's guidelines (para 41), even systems with (limited) inferring capabilities may possibly fall outside the scope of the definition because they are only limitedly able to analyse patterns and independently adapt their outputs. Examples cited include basic data processing systems, systems based on classical heuristics, and simple prediction systems (see paras 42-51 of the Commission guidelines for more details and examples).

[37] C(2025) 5053 final para 18.
[38] C(2025) 5053 final para 17.
[39] Recital 12 AI Act.
[40] See explicitly in C(2025) 5053 final para 23.
[41] Recital 12 AI Act.
[42] See explicitly in C(2025) 5053 final para 26.
[43] Recital 12 AI Act. The European Commission's guidelines provide a more detailed discussion of various AI technologies – see there: C(2025) 5053 final para 32-39.
[44] Recital 12 AI Act.

5. An AI system operates according to **explicit or implicit objectives**.

- ➔ AI systems are designed to operate according to one or more objectives through the fulfilment of tasks. These objectives can be explicit (clearly stated and directly encoded into the system) or implicit (not explicitly stated but may be deduced from behaviour or underlying assumptions of the system).[45]
- ➔ The Commission guidelines cite as a potential objective, for example, answering questions on a set of documents as accurately as possible and with a low rate of failures.[46]

6. The outputs of an AI system may include **predictions, content, recommendations, or decisions**.

- ➔ The AI Act names four categories into which outputs generated by AI systems may fall. The Commission guidelines elaborate on these:
  - **Predictions:** "*estimate about an unknown value (the output) from known values supplied to the system (the input)*".[47]
    - E.g., AI systems for energy consumption forecasts.[48]
  - **Content:** generation of new material – e.g., texts, images, videos, music.[49]
  - **Recommendations:** "*suggestions for specific actions, products, or services to users based on their preferences, behaviours, or other data inputs*".[50]
  - **Decisions:** "*conclusions or choices made by a system*".[51]
- ➔ The Commission guidelines clarify that AI systems can generate such outputs using their capability to handle complex relationships and patterns in data.[52]

[45] C(2025) 5053 final para 24.
[46] C(2025) 5053 final para 25.
[47] C(2025) 5053 final para 54.
[48] C(2025) 5053 final para 55.
[49] C(2025) 5053 final para 56.
[50] C(2025) 5053 final para 57.
[51] C(2025) 5053 final para 58.
[52] C(2025) 5053 final para 59.

7. The outputs of an AI system can **influence physical or virtual environments**.

➔ According to the Commission guidelines, this underlines the fact that AI systems are not passive — they actively impact their environment.[53]

Key Roles of Educational Institutions Under the AI Act

The obligations set out in the AI Act depend on the role an institution or individual plays in relation to an AI system. In the education sector, two roles are particularly relevant:[54]

- provider
- deployer

Definition: A provider of an AI system, according to Article 3(3) of the AI Act, is "*a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge*".

"Placing on the market", according to Article 3(9), refers to the first making available on the Union market, "putting into service", according to Article 3(11), to the supply for first use directly to the deployer or for own use.

[53] C(2025) 5053 final para 60.
[54] In addition to providers and deployers, the AI Act also recognises the roles of authorised representatives, importers and distributors — all of whom have certain obligations as well.

Definition: A deployer of an AI system, under Article 3(4) AI Act is "*a natural or legal person, public authority, agency or other body using an AI system under its authority*".

- **Excluded**: use in the course of a personal non-professional activity

Important: Role transitions in the case of high-risk AI systems possible!

Article 25(1) of the AI Act stipulates that deployers **themselves are considered to be providers of a high-risk AI system** if they:

- **put their name or trademark on a high-risk AI system**,
- **make a substantial modification to a high-risk AI system** in such a way that it **remains a high-risk AI system**, or
- **modify the intended purpose of an initially non-high-risk AI system** in such a way that it thereby **becomes a high-risk AI system**.

In such cases, they must comply with stricter obligations, e.g., to conduct a conformity assessment, create technical documentation, and operate a risk management system.

Recommendation: Particularly in case of technical changes to high-risk AI systems or when changing the intended purpose, carefully check whether this results in a shift from the deployer to the provider status – and which additional obligations this may entail!

Educational institutions **may act as either**, depending on the context.

In most cases, educational institutions will use AI systems provided by third parties (such as commercial software, e.g., for supporting teaching, assessment, or administration). In such cases, the institution is considered a **deployer**.

Example: A school uses an externally provided AI system that automatically analyses essays created by learners according to certain criteria. In this case, the manufacturer of the system is the provider and must meet the relevant provider obligations. The school, as the user of the AI system, acts as a deployer and is subject to fewer requirements.

However, if an institution develops its own AI system, it may qualify as a **provider**. This also applies, as already explained above, in case of **relabelling** or **substantial modification**[55] **of an existing high-risk AI system** or if a **modification of the intended purpose of a non-high-risk AI system leads to its classification as a high-risk AI system**. In such cases, all corresponding provider obligations apply.[56]

Example: A university provides its staff with access to an AI-powered research assistant tool developed by a tech company. In this case, the university acts merely as deployer of the AI system, as it only draws on an existing system. The company that developed and provides it remains the provider.

If, in contrast, the university develops its own AI tool that is subsequently used by its employees, it is itself the provider of the system. This also applies if it does not develop a completely independent AI system but, for instance, substantially modifies an existing high-risk AI system – in this case, it is considered to be the provider and must now also fulfil those obligations laid down in the AI Act for providers.

[55] "Substantial modification" is defined in Article 3(23) as "*a change to an AI system after its placing on the market or putting into service which is not foreseen or planned in the initial conformity assessment carried out by the provider and as a result of which the compliance of the AI system with the requirements set out in Chapter III, Section 2 is affected or results in a modification to the intended purpose for which the AI system has been assessed*".

[56] This is also evident from the "Legal and pedagogical guidelines for the educational use of generative artificial intelligence in the European Schools" issued by the Schola Europaea: these guidelines prohibit staff from actively developing new large language models (LLMs), as doing so would qualify them as providers under the AI Act. This is to be avoided due to the associated regulatory burden. See Schola Europaea, 'Legal and pedagogical guidelines' 7. See also T Hoeren, 'Rechtsgutachten zur Bedeutung der europäischen KI-Verordnung für Hochschulen [Expert Legal Opinion on the Significance of the European AI Act for Higher Education Institutions]' (2025) 17-25 <https://doi.org/10.13154/294-13421> (last accessed 26 August 2025) for a detailed discussion.

Guiding Questions:

- Do we know our role – is the educational institution or an individual staff member acting as a deployer or as a provider?
- Is the AI system merely used by the institution or by staff members, or is it developed under their own name and put into operation/placed on the market?
- Is a high-risk AI system used unchanged or is it adapted? If adapted, are any modifications substantial, so that provider obligations apply?
- Does our use of an AI system that is not considered high-risk change the intended purpose of the system in such a way that it becomes high-risk?

AI Literacy as a Fundamental Requirement

Definition: According to Article 3(56) of the AI Act, AI literacy means "*skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause*".

A key concern of the AI Act is the **promotion of knowledge about AI systems and their responsible use**. Article 4 therefore requires providers and deployers of AI systems to ensure sufficient AI literacy:

- "*Providers and deployers of AI systems shall take measures* ***to ensure, to their best extent****, a* ***sufficient level of AI literacy*** *of* ***their staff and other persons dealing with the operation and use of AI systems on their behalf****, taking into account their* ***technical knowledge****,* ***experience****,* ***education and training*** *and the* ***context*** *the AI systems are to be used in, and considering the* ***persons or groups of persons on whom the AI systems are to be used****.*"

The expression "*their staff and other persons dealing with the operation and use of AI systems on their behalf*" includes not only their own employees, but all "*persons broadly under the organisational remit*", including, for example, external teaching staff engaged on a freelance

basis or external service providers.[57] They must be appropriately trained if they are directly operating an AI system. Providers and deployers – such as an educational institution when it acts in this role – **must ensure this**.[58] In the education sector, this particularly affects:

- leaders responsible for decisions about AI deployment,
- IT and other administrative personnel working with AI systems, and
- educators who use AI systems.

Important: This obligation applies to **all systems falling under the scope of the AI Act**, not only to specific risk categories.

Note: It is **not yet fully clarified** whether the AI Act requires that learners also demonstrate sufficient AI literacy. As end users or affected persons, they are not explicitly mentioned in Article 4. However, where an institution **mandates** that learners use an AI system in order to fulfil academic requirements, such use could arguably be considered "on behalf of" the institution within the meaning of Article 4. Furthermore, recital 20 emphasises the importance of AI literacy for all participants, including affected persons. Educational institutions are therefore advised to promote AI literacy among their students. Additionally, **other legal provisions, regulatory requirements and administrative regulations (including national or regional)** may establish an obligation to provide AI-related training to learners!

For further reading on the topic of advancing AI literacy among young learners, see OECD, in cooperation with the European Commission, *Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education* (Review Draft, 2025) https://ailiteracyframework.org.

[57] See, for instance, European Commission, 'AI Literacy – Questions & Answers' (8 July 2025), <https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers> (last accessed 26 August 2025); see also A Eickmeier and C Petrasch, 'Art. 4 KI-Verordnung: Die unterschätzte Herausforderung auf dem Weg zur KI-Compliance' (YPOG Insights, January 2025) <https://www.ypog.law/insight/art-4-ki-verordnung> (last accessed 26 August 2025).

[58] See, for instance, T Hoeren, 'Rechtsgutachten zur Bedeutung der europäischen KI-Verordnung für Hochschulen [Expert Legal Opinion on the Significance of the European AI Act for Higher Education Institutions]' (2025) 14-15 <https://doi.org/10.13154/294-13421> (last accessed 26 August 2025).

The knowledge and competencies required depend on the context of use and may include:

- **basic technical understanding** of how the relevant AI systems work,
- skills in **interpreting AI-generated results correctly**,
- awareness of the **opportunities, risks, and possible harmful effects**, as well as
- understanding of how AI-influenced decisions **may affect individuals**.[59]

Note: The AI Act does not prescribe how sufficient AI literacy should be achieved – it requires the implementation of **suitable measures**. The European Artificial Intelligence Office provides a continuously updated overview of best practice examples (**Living Repository of AI Literacy Practices**) at the following link:

→ https://digital-strategy.ec.europa.eu/en/library/living-repository-foster-learning-and-exchange-ai-literacy

Replicating the practices listed in this repository does not automatically grant presumption of compliance with Article 4 of the AI Act. **Always check** the relevant legal and administrative guidance (especially from educational authorities) for the appropriate actions to take in the educational sector!

Example: A university lecturer, on behalf of the faculty leadership, uses an AI system to generate exam questions and uses these in the subsequent test. She has received no information about the system's functioning, error risks, or possible consequences. Although the AI-generated questions appear plausible at first glance, it turns out that they contain unclear wording, technical terms are used incorrectly, and the answer options are partly misleading. This causes confusion among students during the exam, and several questions must be discarded.

In this case, the required AI literacy in the exam context is lacking. The uncritical use of an unsuitable AI system leads to poor-quality results, affects the course of the exam, and ultimately the performance assessment. Sufficient AI literacy would have enabled an informed understanding of the system's capabilities and risks.

[59] See also recital 20 AI Act.

Guiding Questions:

- Who in the educational institution is responsible for ensuring AI literacy?
- Which individuals in the educational institution are directly involved in developing, deploying, and/or using AI systems?
- Do these individuals have sufficient AI literacy (e.g., through professional development? Can this be documented?
- Are there training or information programmes available within the educational institution? Are there known external offers that can be recommended?
- Are learners and (if applicable) parents/guardians adequately informed? Is learners' AI literacy considered and supported where necessary?

The Four Risk Categories and Their Requirements

In practice, **four risk categories**, each associated with varying levels of obligations, are distinguished based on the AI Act.[60] **These categories determine key responsibilities of providers and deployers**: the higher the risk, the stricter the requirements. Before using an AI system, it must therefore be clarified which category it belongs to and what this implies for institutions, leaders, and staff in their respective roles.

AI Systems with Minimal or No Risk

For AI systems associated with no or only minimal risk, the AI Act does not introduce specific requirements – **they can be used freely**.[61] Typical examples include AI-supported spam filters or games.[62]

[60] The AI Act explicitly mentions only two categories: prohibited AI practices and high-risk AI systems. However, the remaining two categories are derived from the content of the regulation.
[61] See, for instance, European Commission, 'AI Act' <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (last accessed 26 August 2025); Council of the European Union, 'Artificial intelligence act '. <https://www.consilium.europa.eu/en/policies/artificial-intelligence/> (last accessed 26 August 2025).
[62] For instance, in Council of the European Union, 'Artificial intelligence act <https://www.consilium.europa.eu/en/policies/artificial-intelligence/> (last accessed 26 August 2025).

Example: An AI-supported spam filter in a school's email system automatically filters out advertising emails sent to the administration. This filter can be freely used, and the AI Act imposes no specific rules for its use.

AI Systems with Limited Risk

This category includes **certain AI systems** that **carry some level of risk**, particularly concerning impersonation or deception, irrespective of whether they qualify as high-risk or not.[63] To minimise such risks, the AI Act establishes **transparency obligations**: Users must be informed of the use or that they are interacting with an AI system or about the AI origin of content (text, images, videos, or audio).

The AI Act assigns the following AI systems to this category:

AI systems **intended to interact directly with natural persons**: Article 50(1), see also recital 132

- **Providers** must ensure during the design and development of the AI system that the persons concerned are informed that they are interacting with an AI system.
- **Exceptions**:
  - When it is obvious,[64]
  - In law enforcement contexts.

If the system is intended to interact with persons belonging to **vulnerable groups** due to their age or disability, this must be taken into account![65]

AI systems that **generate synthetic audio, image, video or text content**: Article 50(2), see also recital 133

- **Providers** must ensure that AI-generated content is marked in a machine-readable format and detectable as artificially generated or manipulated. This also applies to general-purpose AI systems.

[63] Recital 132 AI Act.
[64] The standard is the point of view of someone who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use.
[65] Recital 132 AI Act.

- **Exceptions:**
  - When the AI system merely performs an assistive function for standard editing;
  - When the AI system does not substantially alter the input data provided by the deployer or their semantics;
  - In law enforcement contexts.

Emotion recognition systems and **biometric categorisation** systems: Article 50(3), see also recital 132

- **Deployers** must inform exposed persons and process any personal data in accordance with data protection rules (notably the GDPR).
- **Exceptions:**
  - In law enforcement contexts.

AI systems that **generate deep fakes**: Article 50(4), see also recital 134

- **Deployers** of an AI system that generates or manipulates image, audio or video content constituting a deep fake are obliged to disclose this artificial generation or manipulation.
- **Exceptions:**
  - In law enforcement contexts.
- **Restrictions** of this obligation apply to:
  - Content that forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, as well as
  - AI systems that generate texts on matters of public interest to publish them for public information.

Note: If the AI system is classified as high-risk under the AI Act (more information in the next chapter), these transparency obligations apply **in addition** to the requirements imposed on high-risk AI systems (Article 50(6)).

According to Article 50(5), the information must be provided **in a clear and distinguishable manner** and must **be accessible**. It must be provided **at the latest upon first interaction or exposure**.

Examples:

A school's AI-supported chatbot answers learners' questions about a specific subject area. It must be clearly and unambiguously identifiable to learners that they are interacting with an AI system.

- **Important:** When an AI system directly interacts with persons, the particular vulnerability of children and adolescents due to their age must be considered. Information must be age-appropriate.

A university provides learners with automatically generated questions and answers for exam practice. Learners must be able to clearly and unambiguously recognise that the content has been generated or manipulated by an AI system.

Guiding questions to support identification of transparency obligations:

- Are AI systems used in ways that involve direct interaction with users (e.g., learners, educators, administrative staff)?
  - Are users informed accordingly?
  - Is the notice clear, unambiguous, and accessibly available, e.g., also for children/adolescents who interact with the AI system?
- Does the AI generate content — e.g., texts, images, or videos? Is this content clearly marked as AI-generated?
- Are AI systems used that are intended to recognise emotions or biometrically categorise individuals?
  - Are affected persons informed?
- Does the AI generate deep fakes?
  - Is this clearly disclosed?

High-risk AI Systems

Some AI applications are considered high-risk under the AI Act. This category concerns the use of **certain** AI systems in **particularly sensitive areas of life, including education**. The AI Act sets out specific and heightened obligations for these systems.

Particularly relevant provisions for the **education sector** can be found in Annex III of the AI Act.[66] There, **AI systems in the field of education and vocational training** are classified as high-risk if they are intended to be used for one of the following tasks:

- to **determine access or admission** or to **assign persons** to educational and vocational training institutions (para 1(3)(a))
- to **evaluate learning outcomes**, including when those outcomes are used to steer the learning process of persons in educational and vocational training institutions (para 1(3)(b))
- for the purpose of **assessing the appropriate level of education** that a person will receive or will be able to access, in the context of or within educational and vocational training institutions (para 1(3)(c))
- for **monitoring and detecting prohibited behaviour** of students during tests in the context of or within educational and vocational training institutions (para 1(3)(d))

[66] In addition to the AI systems and use cases listed in Annex III, AI systems are also classified as high-risk if they are products — or safety components of such products — that fall under certain EU harmonisation legislation. See in more detail Article 6(1) AI Act.

Note: **Such AI systems are not considered high-risk** if they **do not pose significant risks** to health, safety or fundamental rights, including by not materially influencing the outcome of decision making (Article 6(3)). This is the case when the AI system is intended for one of the following tasks:

- to perform a **narrow procedural task** (Article 6(3), subpara 2(a)) — e.g., detection of duplicates among applications (recital 53)
- to **improve the result of a previously completed human activity** (Article 6(3), subpara 2(b)) — e.g., adjusting the language style of an already written document (recital 53)
- to **detect decision-making patterns or deviations from prior decision-making patterns** and is not meant to replace or influence the previously completed human assessment, without proper human review (Article 6(3), subpara 2(c)) — e.g., subsequent review of whether a teacher has deviated from their grading pattern (recital 53)
- to **perform a preparatory task to an assessment** relevant for the purposes of the use cases listed in Annex III (Article 6(3), subpara 2(d)) — e.g., translation of documents (recital 53)

If a provider believes that these criteria apply to their AI system listed in Annex III and it is therefore not high-risk, they must document their assessment and are subject to registration obligations (Article 6(4)).

Important: A counter-exception applies – if an Annex III use case includes **profiling** of persons, the system is **always** considered to be high-risk (Article 6(3), subpara 3).

The AI Act sets out **enhanced requirements** for high-risk AI systems, and under Article 16(a), **providers** are required to ensure compliance. For example:

- **Comprehensive risk management** throughout the entire lifecycle of the system to identify, assess, and mitigate risks (Article 9, recital 65).
- **Training, validation, and testing data sets** as well as **data governance** must meet detailed requirements (Article 10, recital 67).

- Creation of **technical documentation** that must be kept up to date; enabling **automatic recording of events** (Articles 11 and 12, recital 71).[67]
- **Transparency requirements** and **provision of information** (particularly comprehensive instructions for use) for deployers (Article 13, recital 72).
- Enabling **effective and appropriate human oversight** to prevent or minimise risks to health, safety, or fundamental rights (Article 14, recital 73).
- Appropriate level of **accuracy, robustness, and cybersecurity** of the AI system (Article 15, recital 74-77).

In addition, **providers** must fulfil a number of further obligations set out in Article 16(b)-(l), such as establishing a **quality management system**,[68] keeping **documentation and logs**,[69] fulfilling **registration obligations**, and affixing CE markings[70].

Deployers of high-risk AI systems also have obligations, e.g.:

- **Ensuring proper use:** appropriate technical and organisational measures to ensure use in accordance with the instructions for use (Article 26(1), recital 91).
- **Ensuring human oversight:** Human oversight must be assigned to competent, trained, and authorised persons and these must be supported in this task (Article 26(2), recital 91).
- **Obligations regarding input data:** Insofar as the deployer exercises control over input data, they must ensure that input data is relevant and sufficiently representative in view of the intended purpose of the AI system (Article 26(4)).
- **Monitoring and information obligations in case of incidents:** monitoring the operation of the AI system on the basis of the instructions for use, fulfilling information obligations towards the provider and authorities, e.g., in case of incidents (Article 26(5)).
- **Retention of logs** automatically generated by the AI system (Article 26(6)).
- **Information of affected workers:** These must be informed before the putting into operation or use of the AI system in the workplace (Article 26(7), recital 92).

[67] See for more details on the technical documentation Annex IV to the AI ACt.
[68] See in more detail Article 17 and recital 81 AI Act.
[69] See in more detail Articles 18 and 19 AI Act.
[70] See in more detail Article 48 and recital 129 AI Act.

- **Data protection impact assessment:** Where applicable, deployers must conduct a data protection impact assessment according to data protection provisions (particularly the GDPR) (Article 26(9)).
- **Information of affected persons:** Deployers must inform persons affected by decisions of or assisted by the AI system (Article 26(11)).
- **Fundamental rights impact assessment** (from 2 August 2026): If the deployer is a body governed by public law or a private entity providing public services, conducting a fundamental rights impact assessment may be required before deployment, which must then be communicated to authorities (Article 27). This may particularly include educational institutions.[71]
- **Right to explanation of decision-making:** If a deployer makes decisions on the basis of the output from a high-risk AI system, the persons affected thereby have a right to obtain from the deployer clear and meaningful explanations:
    - of the role of the AI system in the decision-making procedure and
    - of the main elements of the decision taken (Article 86).

  The prerequisite for this right is that the decision produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights.[72]

Example: An AI-based grading tool autonomously assigns final marks for student exams. Since this system is used as intended in the education sector for assessing learning outcomes, it qualifies as a high-risk AI system. However, if an AI system is only used to generate non-binding suggestions that are critically reviewed and confirmed by a person, the exception in Article 6(3) of the AI Act may apply, in which case the system is not considered high-risk.

Guiding questions to support recognition of high-risk AI systems in education:

- Does the AI system influence access or admission to education or does it assign persons to educational institutions?

[71] Compare also recital 96 AI Act.
[72] Compare also recital 171 AI Act.

- Is the AI system involved in assessing individual learning outcomes or educational levels?
- Is the AI system used to monitor students during exams?

Prohibited AI Practices

Some applications of AI systems are **completely prohibited** by the AI Act because they are incompatible with EU fundamental values or pose an unacceptable level of risk. The "prohibited AI practices" particularly include those through which persons can be manipulated or deceived, discriminated against, or impermissibly monitored. **Such AI systems must not be placed on the market, put into service, or used (for this specific purpose).**[73]

Caution is therefore also required in the educational context. The following prohibited AI practices listed in Article 5(1) are particularly relevant for educational institutions and educators:

Emotion recognition in the workplace and in educational institutions: Article 5(1)(f)

- AI systems "*to infer emotions of a natural person in the areas of workplace and education institutions*"
- **Exceptions:** for medical or safety reasons

Manipulative or deceptive systems: Article 5(1)(a)

- Subliminal manipulation: AI system "*that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner*

[73] The European Commission has published accompanying guidelines on prohibited AI practices; see European Commission, 'Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)', C(2025) 5052 final, available at <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act> (last accessed 26 August 2025).

that causes or is reasonably likely to cause that person, another person or group of persons significant harm"

Exploitation of vulnerabilities of a natural person or a specific group of persons: Article 5(1)(b)

- AI system "*that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm*"

Social scoring: Article 5(1)(c)

- AI system "*for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following:*
  - *(i) detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;*
  - *(ii) detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity*"

Example of prohibited AI use in an educational context (based on the example in the chapter on *privacy and data protection* in Part I of the guide): A school wants to enable teachers to adapt their methods and topics in lessons to learners' reactions. For this purpose, it plans to use AI that analyses students' emotions in real time and reports back to the educators.

The proposed AI use is prohibited under Article 5(1)(f) — the AI system must not be used.

Guiding questions to support avoidance of prohibited AI practices:

- Does the AI system recognise or infer emotions?
- Could the AI system influence the behaviour of affected persons without them consciously perceiving it or by deceiving them? Are persons as a result led to make decisions they would not otherwise have made and that harm them or others?
- Does the AI system exploit the specific vulnerability of persons (due to their age, disability, or social or economic situation) to manipulate their behaviour so that they or others are harmed?
- Does the AI system evaluate individuals according to their behaviour, personal or personality characteristics, or classify them accordingly? Does this lead to disadvantage or detriment of these persons?

General-Purpose AI Models

Definitions:

According to Article 3(63) of the AI Act, a general-purpose AI model is one "*that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications*".

A general-purpose AI system is defined in Article 3(66) as an AI system "*which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems*".

Many AI applications are based on models that are not designed for one specific purpose but are **suitable for diverse tasks** (e.g., text generation, creation of program code, or image recognition). Such AI models are often referred to as "General-Purpose AI" or "GPAI" for short.

Examples include LLMs (large language models) such as ChatGPT, Claude, or Gemini.

The AI Act places **specific requirements** on such AI models (Chapter V of the AI Act), particularly **special obligations for providers**, independent of the application context. Providers must, for example,

- create and update **technical documentation** of the model (Article 53(1)(a)),
- **provide information for providers of AI systems** that wish to integrate the AI model (Article 53(1)(b)),
- put in place a policy to comply with EU law on **copyright** and related rights (Article 53(1)(c)),
- and create a **summary about the content used for training** of the AI model (Article 53(1)(d)).

Certain exemptions apply under Article 53(2) for providers of AI models that "*are released under a free and open-source licence that allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, the information on the model architecture, and the information on model usage, are made publicly available*".

Note: The specific obligations for general-purpose AI models also apply when such models are integrated into AI systems or form part of them (see recital 97). **They supplement the four risk categories.** If general-purpose AI models are integrated into AI systems – for instance an LLM integrated into a performance assessment AI system – the risk categories must again be used for assessing the system. The corresponding obligations, such as those for a high-risk AI system, must be fulfilled by providers and deployers. In this context, bear in mind the possibility that a deployer may qualify as a provider if a **modification of the intended purpose of a non-high-risk AI system** leads to its reclassification as a high-risk AI system (see the chapter on Key Roles of Educational Institutions Under the AI Act)!

When general-purpose AI models pose **systemic risk** due to their high-impact capabilities,[74] the AI Act also establishes **additional provider obligations**, in particular to

- **conduct model evaluation** (Article 55(1)(a)),
- **assess and mitigate possible systemic risks** arising from the AI model (Article 55(1)(b)),

[74] See Article 51 of the AI Act for further details on classification.

- track, document, and report **serious incidents and corrective measures** (Article 55(1)(c)),
- and ensure an adequate level of **cybersecurity** (Article 55(1)(d)).

Data Protection Requirements

Introduction

The increasing use of AI applications in the educational sector – both in the classroom and in related contexts – also gives rise to a wide range of data protection challenges.

The use of such systems by educational institutions, educators and learners **may involve the processing of personal data within the meaning of the General Data Protection Regulation (GDPR)**[75] – for example, if personal information is actively entered or if the platform collects and analyses user behaviour. Particularly relevant in this context is the use of large language models, as prompts may contain personal data.

Note: In the context of AI, the term "prompt" refers to an input or instruction used by users to elicit a specific response or output from an AI system. A prompt defines the task to be performed and may include additional information, examples, or context to enable the AI to process the request as accurately as possible.

Against this background, it is essential for educational leaders to be aware of the data protection implications of such technologies in order to ensure that their use complies with applicable legal frameworks.[76]

Key questions to consider from a data protection perspective include:[77]

- Are there procedures in place to restrict data access to only those individuals who require it?
- Is access to learners' data protected, and is the data stored securely and used solely for its intended purposes?

[75] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1. Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (last accessed 26 August 2025).

[76] See also D Thiede, 'ChatGPT und der Datenschutz – eine aktuelle Einschätzung' (March 2023) <https://unterrichten.digital/2023/01/23/chatgpt-datenschutz-unterricht-schule/> (last accessed 26 August 2025).

[77] Compare European Commission, 'Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators' (2022) 21 <https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1> (last accessed 26 August 2025).

- Are there mechanisms in place to ensure that sensitive data remains anonymous?
- Are learners and staff informed about what happens to their data, how it is used, and for what purposes?
- Is there a mechanism that enables educators and school administrators to report privacy or data protection concerns?
- Is the AI system compliant with the General Data Protection Regulation?

Legal Framework

The General Data Protection Regulation (GDPR) constitutes the central and most significant legal source in European data protection law. Adopted in 2016, it has been directly applicable in all EU Member States since **25 May 2018**, following a two-year transitional period.[78] The Regulation aims to harmonise data protection law across the European Union and to adapt it to the demands of the digital society. It serves both to protect the rights of natural persons in relation to the processing of personal data and to ensure the free flow of such data within the internal market.[79]

While the AI Act sets out rules for the development, placing on the market, and use of AI systems, the GDPR lays down rules on the processing of personal data. If an AI system processes personal data, the requirements of both legal instruments must therefore be observed.[80] In this context the AI Act should be understood as a complement to the GDPR, as the latter is not fully equipped to address the challenges arising from the rapid development of AI technology.[81]

The central point of reference in data protection law is the concept of **personal data**:

[78] See European Data Protection Supervisor, 'The History of the General Data Protection Regulation' <https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en> (last accessed 26 August 2025).
[79] Article 1(1) GDPR.
[80] See European Digital Education Hub (EDEH), 'Explainable AI in education: Fostering human oversight and shared responsibility' (2025) 27-28 <https://knowledgeinnovation.eu/kic-publication/explainable-ai-in-education-fostering-human-oversight-and-shared-responsibility/> (last accessed 26 August 2025).
[81] See also D Thiede, 'KI in der Schule – zwischen Datenschutz (DSGVO) und KI-Verordnung (EU AI Act)' (May 2025) <https://unterrichten.digital/2025/05/09/ki-schule-datenschutz-dsgvo-ki-verordnung-eu-ai-act/> (last accessed 26 August 2025).

Definition: According to Article 4(1) of the GDPR, personal data means "*any information relating to an identified or identifiable natural person ('data subject')*". A natural person is considered identified, if they can be clearly determined based on the available information. A person is considered identifiable under Article 4(1) GDPR if their identity can be established by linking the data to additional information.

Example: Personal data may include the following types of information: name, address, photographic images, email address, location data, IP address, device identifiers, and biometric data.

Another key concept in this context is also the term **processing**:

Definition: The term processing is broadly defined in Article 4(2) of the GDPR as "*any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means*". This includes, for example, the collection, storage, alteration, retrieval, use, transmission, dissemination, alignment, combination, or erasure.

Example: A school uses an AI system to administer a test in a way that initially does not record learners' names. However, once the test results are generated, they are assigned to individual learners, listed, and stored.

In this case, personal data is being processed because the learners can ultimately be identified. Although the data collection may appear anonymous at first, the subsequent linkage to individuals means the data is only pseudonymised, not anonymised. The processing therefore falls under the scope of the GDPR.

The GDPR applies to the processing of personal data by **controllers** and **processors**. Specifically, it applies when (a) the processing is carried out **in the context of the activities of an EU establishment** (Article 3(1)); or (b) a **non-EU controller or processor** processes personal data related to **offering goods or services to individuals in the EU** or to the **monitoring of their behaviour in the EU** (Article 3(2)).

Definition: According to Article 4(7) of the GDPR, a controller is "*the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data*".

Definition: According to Article 4(8) of the GDPR, processor means "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*".

Depending on the task performed and the organisational structure of the educational sector in each country, educational Institutions can act as data controllers as well as data processors, although they will, in most cases, act as data **controllers**.[82]

When processing personal data, the following **principles** must be observed in accordance with Article 5 of the GDPR:

- Lawfulness, fairness, transparency (Article 5(1)(a))
- Purpose limitation: Data must be collected for specified, explicit, and legitimate purposes only and not further processed in a manner that is incompatible with those purposes (Article 5(1)(b))
- Data minimisation: Only data that is necessary may be processed (Article 5(1)(c))
- Accuracy: Data must be accurate and kept up to date (Article 5(1)(d))
- Storage limitation: Data may be stored only for as long as necessary (Article 5(1)(e)
- Integrity and confidentiality: Ensure appropriate security of data, including protection against unauthorised or unlawful processing and accidental loss, destruction or damage (Article 5(1)(f))

According to Article 5(2), the **controller** is responsible for compliance with these principles and must be able to demonstrate such compliance (accountability).

[82] Most GDPR obligations rest with the controller. Under Article 24(1), the controller must "implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with" the regulation. Article 25 further promotes data protection by design and by default. By contrast, processors mainly have duties relating to security, record-keeping, and cooperation – see in particular Articles 28-33.
For more information on the distinction of the controller and processor role, see also, for instance, European Data Protection Board, 'FAQ – SME Data Protection Guide' <https://www.edpb.europa.eu/sme-data-protection-guide/faq-frequently-asked-questions_en> (last accessed 26 August 2025).

For processing to be **lawful**, one of the conditions listed in Article 6(1) GDPR must be met:

- consent of the data subject (a)
- performance of a contract (b)
- compliance with a legal obligation (c)
- protection of vital interests (d)
- performance of a task carried out in the public interest or in the exercise of official authority (e)
- legitimate interests pursued by the controller or a third party (f)

Example: A school is planning to use an AI system for individual learning analysis that evaluates students' personal performance data. For this purpose, the students concerned must consent to the processing of their data. However, the question arises as to whether such consent can be considered valid in the educational context. For consent to be lawful under Article 6(1)(a) in conjunction with Article 4(11) of the GDPR, consent must be freely given, informed, and unambiguous. In school settings, this can be problematic, as there is often a structural relationship of dependency between students and the educational institution, and peer pressure within the class may affect the voluntariness of consent. Therefore, in many cases, it may be doubted whether a truly free choice exists. A case-by-case assessment is recommended.

The following **requirements** must also be observed:

- The **controller** must comply with the **information obligations** set out in Articles 12 to 14 of the GDPR.
- Data subjects are entitled to the rights laid down in Articles 15 to 22 of the GDPR, in particular the **right of access, erasure, and objection**. It is the responsibility of the **controller** to facilitate these rights.

Particularly relevant in the context of AI systems is Article 22 of the GDPR, which **protects data subjects against decisions based solely on automated processing** – including profiling. Such decisions are only permitted by way of exception under Article 22(2),[83] if they

- are necessary for entering into or performing a contract,

[83] See also recital 71 GDPR.

- are authorised by Union or Member State law that provides appropriate safeguards for the rights and freedoms and legitimate interests of the data subject,
- or are based on the data subject's explicit consent.

Common Practical Challenges

Many AI systems – especially generative ones – raise significant concerns from a data protection perspective. This chapter highlights practical issues that frequently arise in this context.

Server locations and data linking

- Many generative AI services are operated on servers located **outside the EU**.
- Prompts entered during use are often utilised for the **ongoing training** of the underlying AI model.
- In many cases, the use of such systems requires the creation of a user account (e.g., by providing an email address). There is a risk that login data may be linked to data generated during use. This is particularly problematic when the input contains additional personal data. Further risks arise from the processing of metadata such as device identifiers, location data, and information collected through cookies and trackers, all of which may enable the identification of the data subject.[84]

Use by Learners

- Risk of **disclosing personal data** when using AI systems in educational settings.
- **Limited control** by teachers over students' prompt inputs.[85]

[84] K Scheiter, E Bauer, Y Omarchevska, C Schumacher und M Sailer, 'Künstliche Intelligenz in der Schule: Eine Handreichung zum Stand in Wissenschaft und Praxis' (2025) 8 <https://www.empirische-bildungsforschung-bmbf.de/img/KI_Review.pdf> (last accessed 26 August 2025).

[85] See also D Thiede, 'ChatGPT & Datenschutz – Update für Schule und Unterricht: März 2024' <https://unterrichten.digital/2024/02/28/chatgpt-datenschutz-unterricht-schule-2024/> (last accessed 26 August 2025).

Examples:

- As part of a homework assignment, learners use AI systems for tasks such as text generation or improving a résumé. In doing so, there is a risk that they may enter personal data into the system.
- As part of an art project, learners are instructed to use a multimodal AI system capable of generating and analysing images. In this context, one of them uploads a photo showing himself together with several clearly identifiable friends. As previously mentioned, image data can also constitute personal data.

Use by Educators

- **In most cases unproblematic under data protection law:** the use of AI to create teaching materials, provided that no personal data is processed.
- By contrast, the use of AI to support post-lesson activities or the evaluation of student performance **may raise data protection concerns**. This is especially the case when AI systems are used to provide individual feedback on student work or to assist in grading. In such instances, the automated processing may involve personal data and thus fall within the scope of data protection law.[86]

Recommendations for Data Protection-Compliant Use of AI

Before implementing AI systems, it is highly recommended that educational institutions **establish binding usage policies** as well as **school-wide strategies and procedures**, which also include measures to ensure compliance with the GDPR.[87]

To address the data protection issues described above, the following recommendations should in particular be taken into account:[88]

[86] See also D Thiede, 'ChatGPT & Datenschutz – Update für Schule und Unterricht: März 2024' <https://unterrichten.digital/2024/02/28/chatgpt-datenschutz-unterricht-schule-2024/> (last accessed 26 August 2025); and D Thiede, 'KI in der Schule – zwischen Datenschutz (DSGVO) und KI-Verordnung (EU AI Act)' (May 2025) <https://unterrichten.digital/2025/05/09/ki-schule-datenschutz-dsgvo-ki-verordnung-eu-ai-act/> (last accessed 26 August 2025).

[87] European Commission, 'Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators' (2022) 26 <https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1> (last accessed 26 August 2025).

[88] Compare D Thiede, 'KI in der Schule – zwischen Datenschutz (DSGVO) und KI-Verordnung (EU AI Act)' (May 2025) <https://unterrichten.digital/2025/05/09/ki-schule-datenschutz-dsgvo-ki-verordnung-eu-ai-act/> (last accessed 26 August 2025).

- ➔ To minimise risks when using AI systems, especially by students, a **non-personalised device** should be used.[89]
- ➔ Students should be **thoroughly informed** about how AI systems work and the potential data protection risks involved **before** using them in the classroom.
- ➔ **Direct interaction** between learners and AI systems operated by large non-European companies involves significant data protection risks. These providers may gain access not only to the content of the interaction but also to associated metadata, such as device identifiers or location information.
    - A more **privacy-friendly** approach is the indirect use of AI systems through the educator (without direct interaction between learners and the platform) or via third-party access provided by the educational institution using anonymised accounts. The use of locally operated AI applications, where no data is transmitted to third parties, also reduces data protection risks.
- ➔ Data entered in prompts **must not be used for training** the underlying AI model.
- ➔ When interacting with AI systems, learners must **not enter any personal information about themselves or other individuals** into prompts. This must also be taken into account by educators if they use AI systems to analyse or evaluate learners' work.
- ➔ When using **multimodal AI systems** – that is, systems capable of processing photos, videos, and audio data as part of prompts – it is essential to ensure that the media used **does not contain any personal data**.
- ➔ Educators should be given the opportunity to conduct **random checks** of learners' interactions with AI systems **after the fact**.
- ➔ Article 22 of the GDPR grants data subjects – in this case, for example, learners – the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning them or similarly significantly affects them.

[89] See also D Thiede, 'ChatGPT und der Datenschutz – eine aktuelle Einschätzung' (March 2023) <https://unterrichten.digital/2023/01/23/chatgpt-datenschutz-unterricht-schule/> (last accessed 26 August 2025).

Example: Use of Automatic Assessment of Student Work

- **Permissible**: A university lector uses an AI system to perform a preliminary analysis of students' written work. Based on predefined criteria, the AI generates an initial assessment. The teacher then reviews the automatically generated results and independently determines the final grade.
- **Not Permissible**: It would be problematic if the lector were to adopt the grades generated by the AI without any human review and directly use them in the final performance evaluation.

Tip: There are now several ways to make the use of large non-European AI systems more privacy-friendly. Some providers reduce the risks associated with direct use by acting as intermediaries, foregoing personalised user accounts, anonymising interactions, and thereby preventing the identification of users through access or metadata. Moreover, an increasing number of EU-based providers are offering their own AI systems, often built on open-source language models.[90] This also applies to specialised platforms designed to automatically assess texts from a large number of students. Even when using a privacy-friendly evaluation platform, it is crucial to ensure that no personal content – especially names, biographical details, or other identifiable information – is entered into the prompts.[91]

[90] See also D Thiede, 'KI in der Schule – zwischen Datenschutz (DSGVO) und KI-Verordnung (EU AI Act)' (May 2025) <https://unterrichten.digital/2025/05/09/ki-schule-datenschutz-dsgvo-ki-verordnung-eu-ai-act/> (last accessed 26 August 2025).

[91] See also D Thiede, 'KI in der Schule – zwischen Datenschutz (DSGVO) und KI-Verordnung (EU AI Act)' (May 2025) <https://unterrichten.digital/2025/05/09/ki-schule-datenschutz-dsgvo-ki-verordnung-eu-ai-act/> (last accessed 26 August 2025).

Children's Rights and the Use of AI in Education

The **United Nations Convention on the Rights of the Child** (UN CRC) is an international treaty that grants children (defined as persons aged 0 to 18) their own set of rights. In addition to most countries around the world, the EU is also a party to the UN CRC.[92] Moreover, **Article 24 of the Charter of Fundamental Rights of the European Union** (CFR) incorporates children's rights into EU primary law. Many states have likewise integrated children's rights into their national legal systems.[93]

Children's rights place the child and their needs at the centre and are based on the understanding that children are not merely objects of protection, but **independent rights-holders and actors in their own right**.[94]

Article 24 CFR states:

1. *"Children shall have the right to such protection and care as is necessary for their well-being. They may express their views freely. Such views shall be taken into consideration on matters which concern them in accordance with their age and maturity."*
2. *"In all actions relating to children, whether taken by public authorities or private institutions, the child's best interests must be a primary consideration."*
3. *"Every child shall have the right to maintain on a regular basis a personal relationship and direct contact with both his or her parents, unless that is contrary to his or her interests."*

From a legal perspective, both the AI Act and the GDPR must be **interpreted in light of the CFR**. This means they must be applied and interpreted in a way that safeguards Article 24 CFR. At the same time, its provisions must be balanced against other fundamental rights. Neither the AI Act nor the GDPR may conflict with primary EU law — and therefore not with Article 24 CFR.

For educational institutions, educators, and parents or guardians, this implies the following:

[92] M Bertel, 'BVG Kinderrechte – Vorbemerkungen' in Korinek, Holoubek, Bezemek, Fuchs, Martin and Zellenberg (eds), Österreichisches Bundesverfassungsrecht (18. Lfg, 2023) 3. The text of Article 24 CFR is available at: <https://fra.europa.eu/en/eu-charter/article/24-rights-child> (last accessed 26 August 2025).
[93] An overview of these provisions is available at the following link: <https://fra.europa.eu/de/eu-charter/article/24-rechte-des-kindes#national-constitutional-law> (last accessed 26 August 2025).
[94] M Bertel, 'BVG Kinderrechte – Vorbemerkungen' in Korinek, Holoubek, Bezemek, Fuchs, Martin and Zellenberg (eds), Österreichisches Bundesverfassungsrecht (18. Lfg, 2023) 8.

- **Children's rights should be taken into account** in the development and training of AI systems.
- Whenever possible, **AI systems should be selected that uphold children's rights**.
- When applying the AI Act or the GDPR, **Article 24 CFR should be considered in any matter involving children** — particularly regarding the child's best interests and their right to be heard.

Key questions from a children's rights perspective include:

- Is the use of the AI system in line with the child's best interests?
- Were children's rights taken into account in the development of the AI system?
- Have risks specific to children (e.g., emotional manipulation) been assessed and mitigated?
- Are there mechanisms that enable the child's views to be considered in decisions about AI use?
- Is the child, where appropriate, informed in a clear and understandable way about the AI system and his or her rights?
- Is there a procedure for children or their guardian to raise concerns or objections regarding the use of AI systems?

Illustration: Key Legal Aspects at a Glance

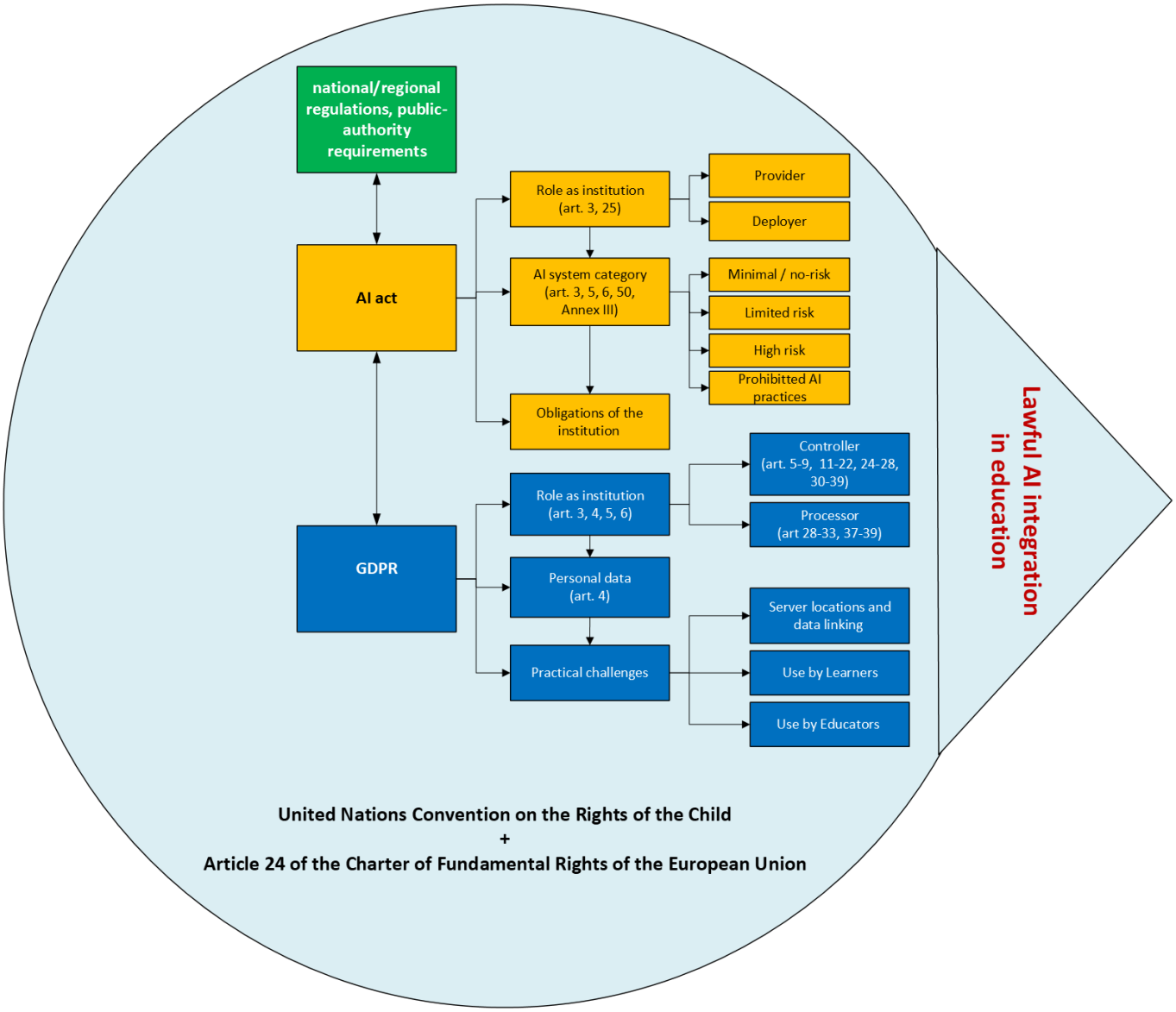

Illustration 2: Key Legal Considerations for AI Integration in Education

This illustration highlights essential legal considerations for integrating AI at both the curricular and institutional level. Educational leaders (both educators and managers) should keep in mind the main legal frameworks, such as the AI Act, the GDPR, and national or regional rules, along with any requirements from public authorities. This overview focuses on selected aspects of compliance and does not replace a full legal review.

Conclusion

The integration of AI into educational settings offers great potential – but also raises complex legal and ethical challenges. This guide is intended to support educational institutions in navigating these challenges, particularly with regard to data protection, children's rights, and the obligations under the AI Act. Given the rapidly evolving legal landscape and the particular responsibilities that come with using AI systems in environments involving children and adolescents, a careful and well-informed approach is essential. By promoting transparency, fairness, and AI competence, and by choosing appropriate systems and practices, educational leaders can help ensure that AI contributes meaningfully and responsibly to teaching and learning. Continued attention to legal developments and ethical principles will remain key.

Annex: The Guide in a Nutshell

This summary offers a concise overview of the key points from the guide, highlighting essential considerations for using AI in education. It includes guiding questions to help assess whether a specific AI system or its intended use might raise legal, ethical, or practical concerns.

However, this summary is not a substitute for a comprehensive evaluation and full compliance check of the AI system!

Also take into account that the focus of the guide is on the AI act as well as the GDPR. Other legal acts (EU as well as national legal acts) are not taken into account, but might apply to your case.

Ethical AI Integration in Education

Essential Aspects

The following aspects are key to guiding the ethical integration of AI in education.

Placing People at the Core of AI Ethics

At the heart of AI ethics is a human-centred approach: the use of AI systems should be guided by the well-being and empowerment of the individuals affected.

The Importance of Context in AI Use

AI technologies and their potential applications in education are highly diverse. Which ethical considerations are most relevant depends both on the AI system's capabilities and functionality, but also on the specific use case.

Embedding Ethics Across the AI Lifecycle

Ethical principles should be considered from the outset and throughout the entire lifecycle of an AI system: from its initial design and technical development to its practical implementation.

Step 1: Before Deployment

Assessing Risks

A careful risk assessment must be carried out before any AI system is deployed. This is particularly important in schools, where trust and protection play a central role.

- Is there a clearly designated individual or team responsible for assessing potential risks prior to AI use?
- Has this assessment been carried out and documented? Were social impacts and specific vulnerabilities (e.g., age, emotional development) taken into account?

Human Autonomy

- Is individual decision-making of all stakeholders appropriately taken into account — e.g., by offering options or alternative approaches?
- Is there a risk that individuals place too much trust in the AI system (e.g., uncritically following AI-generated suggestions) or become dependent on it?

Respect of needs

- Has the AI system been tested and found suitable for use in education and for the relevant age group?
- Is it ensured that contents are age-appropriate, particularly in interactive AI systems?
- Do teachers, pupils and administrative staff need additional training on the AI system that will be deployed, to ensure it is used in an informed and responsible way?

Fairness, Inclusion, Equal Opportunities, and Non-Discrimination

- Is it ensured that the AI system does not contain or reproduce biases?
- Is the system applied fairly to all individuals (e.g., not disproportionately to already disadvantaged groups)?

Practical advice: Based on the risk assessment and taking into account all stakeholders, including children and young people, establish **institutional guidelines** for the use of AI. This will ensure greater security and transparency for all.

Step 2: While Deploying

Ensuring Human Oversight in AI Decisions

Essential decisions must not be left entirely to automated processes. Oversight must always remain with humans.

- Which oversight model is appropriate in light of these risks?
- At which stages is human oversight required — e.g., at the start, at intervals, continuously?

Explainability and Transparency in AI Systems

- Are the basic methods, functions, and decision-making mechanisms of the AI system known? Are they (at least in simplified form) explainable to affected persons?
- Is it clear to those affected when they are interacting with an AI system (e.g., with chatbots)?

Engaging Stakeholders in AI Use

- Can all affected stakeholders (e.g., learners, educators, administrative staff) participate in shaping the use of the AI system or choose alternatives?

Safeguarding Privacy and Protecting Data

- What data does the AI system collect or process? Does this include personal or particularly sensitive data?
- Is it possible to anonymise or pseudonymise the data?

Step 3: After Deployment

Establishing Accountability and Complaint Mechanisms

Even when technology is used to support or automate processes, human responsibility for the design, selection, and application of AI systems must be maintained.

- Are there clearly named individuals within the institution responsible for selecting, implementing, and monitoring AI systems? Do they have sufficient expertise in ethical, pedagogical, and legal matters?
- Are there clear rules on how much influence AI-generated results can have on human decisions (e.g., performance assessments)?
- Do all involved understand that they remain responsible for their decisions – even when supported by AI?

Lawful AI Integration in Education

Understanding the Legal Context of AI Use

- Legal obligations vary depending on your role, the AI system you are using, and its purpose.
- In most cases, multiple legal frameworks will apply in a given context. Therefore, you might have to follow different legal norms – EU law alongside national rules that may vary across countries.
- Legal requirements also evolve over time. Always make sure you rely on the most recent laws, administrative regulations, and policy guidelines.

Step 1: Compliance with EU Law

Verify that all AI systems used in education meet the requirements of the AI Act, including risk classification and transparency obligations, as well as GDPR standards for data protection, consent, and privacy.

AI Act

1) Know Your Role:

- Who is acting: the educational institution or the individual staff member?
- Are you (the educational institution, individual staff member) a provider or a deployer?
    - If you are further developing the AI system, you could be a provider bearing more obligations than a deployer.
    - If you are only using the AI system, you are only bound by the obligations for deployers.

2) Know the AI System You are Using:

Four risk categories are commonly distinguished based on the provisions of the AI Act: minimal/no risk, limited risk, high-risk, prohibited AI practices.

- Does the AI system interact directly with users or generate content (including deep fakes)? The AI system you are using could be a limited-risk AI system, subject to transparency obligations.
- Does the AI system influence access or admission to educational institutions, assess individual learning outcomes, or monitor students during exams? The AI system you are using could be a high-risk AI system, with specific and increased legal obligations.
- Does the AI system recognise emotions in educational institutions, influence behaviour in unnoticed or deceptive ways leading to harmful decisions, or exploit specific vulnerabilities to manipulate behaviour? The use of such an AI system is prohibited.

3) Know Your Obligations

- Using a minimal/no risk AI system (e.g., a spam filter) does not come with specific obligations.
- Using a limited-risk AI system comes with transparency requirements: Users must be informed of the AI use or that they are interacting with an AI system (e.g., Chatbots) or about the AI origin of content (text, images, videos, or audio).
- Using a high-risk AI system comes with specific and heightened obligations for these systems.

4) Know Your Institution:

- Who within the institution is responsible for promoting and ensuring AI literacy?
- Who within your institution decides on the use of AI?
- Who within your institution decides on which AI system to use?

GDPR

1) Know Your Role:

- Role as institution (see Art 3, 4, 5, 6): Is the educational institution acting as a data controller or as a data processor?
- Which principles must be observed when processing personal data, and when can such processing be considered lawful?

2) Know Your Data:

- Are you dealing with any information relating to an identified or identifiable natural person ('personal data')?

3) Know Your System:

- Do you know where the server of generative AI services is hosted? Is it hosted within or outside the EU?
- Is it possible to choose a system that works with servers within the EU?

4) Know Your People:

- Do your students know about data protection when they use AI system for their course work or post-lesson activities?

United Nations Convention on the Rights of the Child and Article 24 of the Charter of Fundamental Rights of the European Union

1) Know Your Students:

- Are you working with children / young people aged between 0 to 18?

2) Know Your System:

- Do you know whether children's rights were taken into account in the development of the AI system?

3) Know the Reasons Why You are Considering an AI System:

- Reflect the use of the AI system: Is its use in line with the child´s / the children's best interests?
- What does the child/do the children think about the use of the AI system? Do they have a choice not to use the AI system?

Step 2: Compliance with National and Regional Legal Requirements

Identify and comply with specific national or regional regulations, such as additional data protection laws, educational standards, or AI-specific rules that may impose stricter obligations than EU legislation. There might also be specific acts on education.

Step 3: Compliance with Requirements of Public Authorities

Proactively engage with relevant public authorities, such as education ministries or data protection agencies, to obtain necessary approvals, adhere to local guidelines, and ensure the AI system aligns with public policies and ethical standards.

References

Bertel, M. (2023). BVG Kinderrechte – Vorbemerkungen. In Korinek, M., Holoubek, M., Bezemek, C., Fuchs, A., Martin, M., & Zellenberg, F. (Hrsg.), *Österreichisches Bundesverfassungsrecht* (18. Lfg 2023).

Deutscher Ethikrat [German Ethics Council]. (2023). *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz; Stellungnahme. [Humans and Machines – Challenges Posed by Artificial Intelligence; Opinion]*. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf.

Eickmeier, A., & Petrasch, C. (2025). *Art. 4 KI-Verordnung: Die unterschätzte Herausforderung auf dem Weg zur KI-Compliance.* YPOG Insights. https://www.ypog.law/insight/art-4-ki-verordnung.

European Data Protection Supervisor. *The History of the General Data Protection Regulation*. https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en.

European Commission. (2025). *AI Literacy – Questions & Answers*. https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers.

European Commission: European Education and Culture Executive Agency. (2023). *AI report: by the European Digital Education Hub's Squad on artificial intelligence in education*. https://data.europa.eu/doi/10.2797/828281.

European Commission. (2025). *Commission guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)* (C(2025) 5053 final). https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application.

European Commission. (2025). *Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)* (C(2025) 5052 final). https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

European Commission. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1.

European Commission. *AI Act*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

European Data Protection Board. *FAQ – SME Data Protection Guide*. https://www.edpb.europa.eu/sme-data-protection-guide/faq-frequently-asked-questions_en.

European Parliament, Council and Commission. (2015). *Joint practical guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation* (2nd edition). https://op.europa.eu/en/publication-detail/-/publication/3879747d-7a3c-411b-a3a0-55c14e2ba732.

European Digital Education Hub. (2025). *Explainable AI in education: Fostering human oversight and shared responsibility*. https://knowledgeinnovation.eu/kic-publication/explainable-ai-in-education-fostering-human-oversight-and-shared-responsibility/.

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. (April 2019). https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Hoeren, T. (2025). *Rechtsgutachten zur Bedeutung der europäischen KI-Verordnung für Hochschulen. [Expert Legal Opinion on the Significance of the European AI Act for Higher Education Institutions].* https://doi.org/10.13154/294-13421.

Kline, R. (2011). *Cybernetics, automata studies, and the Dartmouth conference on artificial intelligence. IEEE Annals of the History of Computing*, 33(4), 5–16. https://doi.org/10.1109/MAHC.2010.44.

Marcelin, T., & Killmayer, L. (2025, June). *AI Act implementation timeline* (At a Glance No. PE 772.906). https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/772906/EPRS_ATA(2025)772906_EN.pdf.

Methnani, L., Tubella, A., Dignum, V., & Theodorou, A. (2021). Let me take over: Variable autonomy for meaningful human control. *Frontiers in Artificial Intelligence, 4*, Article 737072**.** https://doi.org/10.3389/frai.2021.737072.

Müller, V. C. (2025). *Ethics of artificial intelligence and robotics.* In E. N. Zalta & U. Nodelman (eds), *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/ethics-ai/.

OECD. (2025). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education* (Review draft). https://ailiteracyframework.org.

Perković, G., Drobnjak, A., & Botički, I. (2024). Hallucinations in LLMs: Understanding and Addressing Challenges. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, 2084–2088. IEEE. https://doi.org/10.1109/MIPRO60963.2024.10569238.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535.

UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

Council of the European Union. *Artificial intelligence act*. https://www.consilium.europa.eu/en/policies/artificial-intelligence/.

Scheiter, K., Bauer, E., Omarchevska, Y., Schumacher, C., & Sailer, M. (2025). *Künstliche Intelligenz in der Schule: Eine Handreichung zum Stand in Wissenschaft und Praxis*. https://www.empirische-bildungsforschung-bmbf.de/img/KI_Review.pdf.

Schola Europaea, Office of the Secretary-General, Pedagogical Development Unit. (2025). *Legal and pedagogical guidelines for the educational use of generative artificial intelligence in the European Schools* (Ref. 2025-01-D-66-en-2). https://www.eursc.eu/BasicTexts/2025-01-D-66-en-2.pdf.

Thiede, D. (2023). *ChatGPT und der Datenschutz – eine aktuelle Einschätzung*. Unterrichten.digital. https://unterrichten.digital/2023/01/23/chatgpt-datenschutz-unterricht-schule/.

Thiede, D. (2024). *ChatGPT & Datenschutz – Update für Schule und Unterricht: März 2024*. Unterrichten.digital. https://unterrichten.digital/2024/02/28/chatgpt-datenschutz-unterricht-schule-2024/.

Thiede, D. (2025). *KI in der Schule – zwischen Datenschutz (DSGVO) und KI-Verordnung (EU AI Act)*. Unterrichten.digital. https://unterrichten.digital/2025/05/09/ki-schule-datenschutz-dsgvo-ki-verordnung-eu-ai-act/.

Note: The preparation and translation of this guide were supported by the use of AI tools.